Reasoning about cell dynamics using network models

Jacobsen, A.

2019

document version
Publisher's PDF, also known as Version of record

VRIJE UNIVERSITEIT

Reasoning about cell dynamics using network models

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Bètawetenschappen
op dinsdag 29 januari 2019 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door
Annika Jacobsen
geboren te Tórshavn, Faeröer

promotor: prof.dr. J. Heringa
copromotor: dr.ir. K.A. Feenstra

Printed by: Print Service Ede - www.proefschriftenprinten.nl

ISBN: 978-94-92679-74-1

Contents

CHAPTER 1

Introduction

Complex diseases, such as cancer, diabetes and Alzheimer's, have a vast heterogeneous genetic component causing deregulation of cellular systems. The key elements of these systems are molecules, the interactions between them, and the functions that they regulate. Consequently, investigating how these elements differ between healthy and diseased systems provides insights into the cause of a disease, factors driving its progression and possible therapeutic avenues. The molecular interactions of a cellular system can be represented as a network, which enables analysis and reasoning of underlying regulatory mechanisms.

The goal of this thesis is to explore mechanisms of altered molecular networks using computational modeling and analysis in combination with biological experiments, and to develop computational frameworks to enable this. In this Chapter we describe deregulated molecular networks, reasoning about molecular networks, and finally, the outline of this thesis.

1.1 Deregulated molecular networks

Earlier studies of cellular molecules and their roles in cellular functions were predominately based on reductionisms, i.e. each molecule was considered individually (Singh, 2003). This was a highly successful approach for decades and resulted in the discovery of crucial cellular molecules and their properties; in particular the role of molecules in heritable diseases such as the *CFTR* gene for cystic fibrosis (Kerem et al., 1989) and the *APC* gene for colorectal cancer (Groden et al., 1991). However, as cellular molecules are interconnected and work together to regulate cellular dynamics many diseases cannot be characterized based on individual molecules, but may be understood at a systems level.

Networks are abstract representations commonly used to describe interactions between cellular molecules. There are different types of interconnected molecular networks, such as gene regulatory, protein-protein interaction, signaling, metabolic, and genetic interaction networks that interact in complex ways. With our current knowledge of the complexity of molecular biology, it is now clear that the flow of information described by the central dogma of molecular biology (Crick, 1970) – DNA to RNA, and RNA to protein – has associated complex regulatory networks. Transcription factors regulate the transcription of genes, and are themselves regulated by cell signaling. Notably, the regulation of transcription factors is controlled by negative feedback loops, which are central to achieving homeostasis in a cell. Multiple signaling pathways can regulate a single transcription factor, but a single signaling pathway can also regulate multiple transcription factors. Similarly,

multiple transcription factors can regulate a single gene, but a single transcription factor can also regulate multiple genes. This underlines the complexity of these networks that are capable of precisely regulating cellular functions due to their highly structural organization. Therefore, molecular alterations may lead to deregulation of cellular functions, which can have severe consequences. In fact, multiple distinct alterations may cause deregulation of the same functions; this can be seen in cancer and other complex diseases having strong heterogeneous genetic components.

1.1.2 Deregulated signaling pathways in colorectal cancer

Colorectal cancer is the third most common cancer in men and the second most common cancer in women worldwide (Ferlay et al., 2015). The primary event in development of colorectal cancer is a sustained proliferative signal caused by deregulated Wnt/β-catenin and Ras-MAPK signaling (Hanahan and Weinberg, 2011), which is largely caused by alterations of genes in the respective pathways. Mutations in genes in the Wnt/β-catenin signaling pathway can lead to progression of normal epithelial cells to adenomas (Fearon, 2011). Further mutations in genes in the Ras-MAPK signaling pathway and other crucial pathways, such as TP53 signaling, drive adenoma-to-carcinoma progression (The Cancer Genome Atlas Network, 2012; Haan et al., 2014). Moreover, chromosomal aberrations occur in ~85% of CRC (Lengauer et al., 1997) and play a crucial role in carcinogenesis (Matano et al., 2015).

The identification of frequently deregulated genes and altered molecular mechanisms in cancers is crucial to better understand carcinogenesis. Driver genes, in particular, can be identified by that they are mutually exclusive, i.e. found mutated together with a lower frequency than expected. The mechanism of mutual exclusiveness can be explained by genetic interactions, which are another crucial characteristics of complex diseases (Vogelstein and Kinzler, 2004; Ciriello et al., 2012; Canisius et al., 2016).

1.1.3 Genetic interaction networks

Genetic interaction (or epistasis) between two genes occurs when a double mutant has an unexpected phenotype compared to the phenotypes of the individual mutations (Phillips, 2008). Genetic interactions have been observed in i.e., functionally related genes in redundant pathways (Hartman et al., 2001) and in genes that take part in the same protein complex or pathway (Collins et al., 2007). Specific genetic interactions patterns, such as buffering, suppression, masking, and inversion, have been observed in gene expression analyses (Sameith et al., 2015). The underlying molecular mechanisms of genetic interaction patterns are, however, not very well understood (van Leeuwen et al., 2017; Lagator et al., 2017; Large et al., 2017). Thus, investigating these is crucial to get a better understanding of the genotype-phenotype relationship in complex diseases.

1.2 Reasoning about molecular networks

To enable reasoning about cellular systems an abstract representation of the biological problem is needed. Here, networks are commonly used consisting of nodes connected by direct or undirected edges. Networks are highly structural and follow basic organizing principles (Barabasi and Oltvai, 2004; Barabasi et al., 2011), and since molecular components often work together to regulate different molecular functions a cellular network can be simplified into modules (parts of the network) of highly interconnected nodes (Hartwell et al., 1999). Each module can then be further simplified into regulatory patterns, called motifs, occurring more frequently than expected (Milo et al., 2002). Although a network representation of a cellular system clearly simplifies the biological reality by not considering spatial and temporal factors and because details of molecular interactions in the network constitute only a part of the system, they do structure our understanding of a system. Additionally, they also capture an important facet of the cellular system: they generally have many inputs and outputs (edges) for each component (node) in the system (network). Thus, a small network representation may provide insights into complex cellular systems. Yet, even for small network representations of cellular systems, it can be difficult to reason about their regulatory mechanisms. But, when the size of the system increases, beyond five elements or so, intuitive reasoning quickly becomes impossible. For instance, answering the question "which connections to known involved factors could be consistent with the observed behavior?" is next to impossible to answer without computational aid. Computational modeling is therefore an important means to study cellular systems and can be used for generating and verifying hypotheses, as well as allowing for qualitative and often even quantitative reasoning at the molecular level.

1.2.1 Computational modeling of molecular networks

Computational modeling approaches can be used as a means for reasoning about the mechanisms and behavior of cellular systems. By creating a model of a cellular system we can, *i*) formally describe our understanding of the system, and thereby *ii*) identify gaps in our knowledge, and *iii*) more easily reason about plausible causative molecular mechanisms. Moreover, we can *iv*) validate our understanding of the system by comparing the model simulation to known behavior of the system, *v*) predict the behavior of the system when e.g., the conditions are infeasible to test experimentally, there are gaps in our current knowledge and we want to understand the unknown variables (hypothesis testing), or all combinations of the variables are tested, and *vi*) provide explanations for systems for which there are competitive hypotheses based on experimental evidence.

To implement computational models of cellular systems four kinds of inputs are needed as specified in Chapter 6: *i*) a definition of the model, *ii*) the parameters and conditions, *iii*) the simulation rules, and *iv*) the rules of analysis. There are different modeling techniques that can be used depending on the problem at hand

(Karlebach and Shamir, 2008). Ordinary differential equations (ODEs) (Goodwin, 1963) are based on fine-grained data, such as protein concentrations and reaction rates, and describe quantitative traits. These models are therefore often required to be small, but offer a high level of detail. Boolean networks (Kauffman, 1969) and Petri nets (Petri, 1962), on the other hand, may be constructed using coarse-grained data, describe qualitative traits, and can therefore be considerably larger than a fine-grained mathematical model.

1.2.2 Petri net modeling

Petri nets can be applied to investigate the dynamics of molecular networks. Petri nets consist of two types of nodes, places and transitions, connected by directed edges. Places can only be connected to transitions and vice versa. Places represent biological entities (e.g. genes and proteins) and transitions represent the activities between them (e.g. gene transcription). The availability of the entities is represented by tokens in their respective places. Tokens are transferred between the places denoting dynamics in the network. The weight on the edge going from the incoming place to the transition denotes the token requirement for consumption. The weight on the edge going from the transition to the outgoing place denotes the number of tokens produced. A transition is enabled if the number of tokens is equal to or higher than the token requirement. The simulation of a Petri net is done in a step-wise manner.

We have applied Petri nets to model biological problems of different networks, from gene regulatory networks to signaling networks, and of different organisms, from *Saccharomyces cerevisiae* to *Caenorhabditi elegans* (Bonzanni et al., 2009; Bonzanni et al., 2009; Bonzanni et al., 2013).

1.3 Outline of this Thesis

In this thesis plausible mechanisms for molecular networks are derived using computational modeling and analysis. We first describe our research on the molecular mechanisms of perturbed signaling pathways in colorectal cancer. Aberrant Wnt/β-catenin signaling is one of the mechanisms driving colorectal tumor progression from normal to adenoma. In Chapter 2, we describe a Petri net model of Wnt/β-catenin signaling, where we simulate the behavior of active and hyperactive signaling with experimental validation, further predicting the hyperactive behavior caused by different oncogenic network perturbations. Progression from adenoma to carcinoma is frequently associated with 20q gain and overexpression of AURKA in colorectal tumors. In Chapter 3, we explore the effect of *AURKA*-gain on Wnt and Ras-MAPK signaling, in order to understand this aspect of the complex interplay between genetic mutations. This is achieved by measuring the effects on AURKA-mediated cellular signaling processes, including how this affects disease progression.

We thereafter describe our studies on examining the underlying mechanisms causing genetic interaction patterns in yeast. In Chapter 4, we perform exhaustive Petri net modeling of genetic interaction patterns, describing the mechanisms of the inversion pattern, which is frequently associated with transcription factors. In Chapter 5, we describe the construction of the framework used for exhaustive Petri net modeling of epistatic patterns (applied in Chapter 4) and describe downstream analyses to propose mechanistic explanations for these patterns.

Finally, in Chapter 6, we describe a framework for automatically generating executable pathway models specified in BioPAX, recapitulating our simulation results from Chapter 2. Finally, a general discussion given the research described in this thesis is presented in Chapter 7.

References

Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011, 12, 56-68.

Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004, 5, 101-13.

Bonzanni N, Feenstra KA, Fokkink W, Krepska E. What can Formal Methods bring to Systems Biology? In: Cavalcanti A, Dams DR (eds) FM 2009: Formal Methods FM 2009 Lecture Notes in Computer Science, vol 5850 Springer, Berlin, Heidelberg 2009a. 2009.

Bonzanni N, Garg A, Feenstra KA, Schutte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013, 29, i80-8.

Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. Bioinformatics. 2009, 25, 2049-56.

Canisius S, Martens JW, Wessels LF. A novel independence test for somatic alterations in cancer shows that biology drives mutual exclusivity but chance explains most co-occurrence. *Genome Biol*. 2016, 17, 261.

Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res*. 2012, 22, 398-406.

Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet*. 2013, 45, 1127-33.

Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*. 2007, 446, 806–10.

Crick F. Central Dogma of Molecular Biology. *Nature*. 1970, 227, 561-3.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015, 136, E359-86.

Goodwin BC. Temporal Organization in Cells; a Dynamic Theory of Cellular Control Process. London: Academic Press. 1963.

Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, et al. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell*. 1991, 66, 589-600.

Haan JC, Labots M, Rausch C, Koopman M, Tol J, Mekenkamp LJ, et al. Genomic landscape of metastatic colorectal cancer. *Nat Commun*. 2014, 5, 5457.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011, 144, 646-74.

Hartman JL, Garvik B, Hartwell L. Principles for the Buffering of Genetic Variation. *Science*. 2001, 291, 1001-4.

Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature*. 1999, 402, C47-52.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013, 502, 333-9.

Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008, 9, 770-80.

Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theoret Biol*. 1969, 22, 437-67.

Kerem B, Rommens J, Buchanan J, Markiewicz D, Cox T, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989, 245, 1073-80.

Kitano H. Computational systems biology. *Nature*. 2002, 420, 206-10.

Lagator M, Paixao T, Barton NH, Bollback JP, Guet CC. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *Elife*. 2017, 6.

Large EE, Padmanabhan R, Watkins KL, Campbell RF, Xu W, McGrath PT. Modeling of a negative feedback mechanism explains antagonistic pleiotropy in reproduction in domesticated Caenorhabditis elegans strains. *PLoS Genet*. 2017, 13, e1006769.

Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature*. 1997, 386, 623–7.

Matano M, Date S, Shimokawa M, Takano A, Fujii M, Ohta Y, et al. Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. *Nat Med*. 2015, 21, 256-62.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. *Science*. 2002, 298, 824-7.

Petri CA. Kommunikation mit Automaten. PhD Thesis, Technische Universität Darmstadt, Bonn, Germany. 1962.

Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008, 9, 855-67.

Sameith K, Amini S, Groot Koerkamp MJ, van Leenen D, Brok M, Brabers N, et al. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biol*. 2015, 13, 112.

Singh RS. Darwin to DNA, molecules to morphology: the end of classical population genetics and the road ahead. *Genome*. 2003, 46, 938-42.

The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012, 487, 330-7.

van Leeuwen J, Pons C, Boone C, Andrews BJ. Mechanisms of suppression: The wiring of genetic resilience. *Bioessays*. 2017, 39.

Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med*. 2004, 10, 789-99.

CHAPTER 2

Construction and experimental validation of a Petri net model of Wnt/β-catenin signaling

PLoS One. 2016, 11(5), e0155743. doi:10.1371/journal.pone.0155743.

Annika Jacobsen, Nika Heijmans, Folkert Verkaar, Martine J. Smit, Jaap Heringa*, Renée van Amerongen*, K. Anton Feenstra*
* shared last authors

Abstract

The Wnt/β-catenin signaling pathway is important for multiple developmental processes and tissue maintenance in adults. Consequently, deregulated signaling is involved in a range of human diseases including cancer and developmental defects. A better understanding of the intricate regulatory mechanism and effect of physiological (active) and pathophysiological (hyperactive) WNT signaling is important for predicting treatment response and developing novel therapies. The constitutively expressed CTNNB1 (commonly and hereafter referred to as β-catenin) is degraded by a destruction complex, composed of amongst others AXIN1 and GSK3. The destruction complex is inhibited during active WNT signaling, leading to β-catenin stabilization and induction of β-catenin/TCF target genes.

In this study we investigated the mechanism and effect of β-catenin stabilization during active and hyperactive WNT signaling in a combined in silico and in vitro approach. We constructed a Petri net model of Wnt/β-catenin signaling including main players from the plasma membrane (WNT ligands and receptors), cytoplasmic effectors and the downstream negative feedback target gene *AXIN2*. We validated that our model can be used to simulate both active (WNT stimulation) and hyperactive (GSK3 inhibition) signaling by comparing our simulation and experimental data. We used this experimentally validated model to get further insights into the effect of the negative feedback regulator AXIN2 upon WNT stimulation and observed an attenuated β-catenin stabilization. We furthermore simulated the effect of APC inactivating mutations, yielding a stabilization of β-catenin levels comparable to the Wnt-pathway activities observed in colorectal and breast cancer.

Our model can be used for further investigation and viable predictions of the role of Wnt/β-catenin signaling in oncogenesis and development.

2.1 Introduction

The Wnt/β-catenin signaling pathway is crucial for regulating cell proliferation and differentiation during embryonic development, while in adults it helps control tissue homeostasis and injury repair in stem cell maintenance (Clevers and Nusse, 2012; Cadigan and Peifer, 2009). Extracellular WNT ligands activate signaling leading to CTNNB1 (commonly and hereafter referred to as β-catenin) stabilization, nuclear translocation, interaction with TCF/LEF transcription factors (Henderson and Fagotto, 2002) and induction of β-catenin/TCF target genes (Mosimann et al., 2009) (Figure 2.1B). A critical feature of Wnt/β-catenin signaling is the inhibition of a 'destruction complex' which degrades the constitutively expressed β-catenin (Figure 2.1A) (Stamos and Weis, 2013).

The destruction complex consists of two scaffolding proteins, AXIN1 and adenomatous polyposis coli (APC), and two kinases, casein kinase 1 (CK1) and glycogen synthase kinase 3 (GSK3). β-catenin is phosphorylated by CK1 and GSK3 (Amit et al., 2002; Ikeda et al., 1998) and thereafter presented to the proteasome

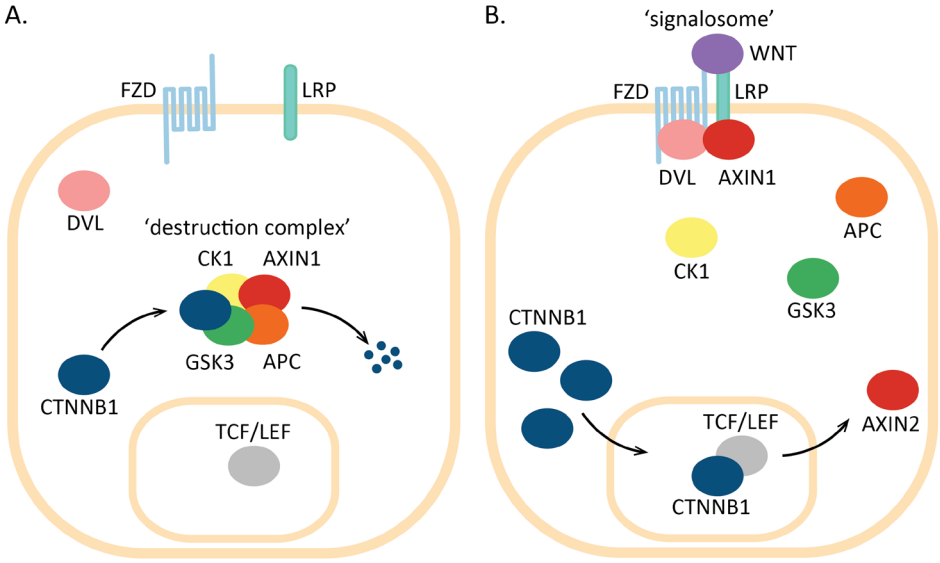


Figure 2.1. Illustration of Wnt/β-catenin signaling. (A) In the absence of an external WNT stimulus β-catenin (referred to by its official gene name CTNNB1 in the figure) is continuously degraded by a 'destruction complex' consisting of AXIN1, adenomatous polyposis coli (APC), casein kinase 1 (CK1) and glycogen synthase kinase 3 (GSK3). (B) Extracellular WNT interacts with the membrane-bound receptors frizzled (FZD) and lipoprotein receptor-related protein (LRP). Dishevelled (DVL) interacts with the intracellular tail of FZD and sequesters AXIN1 to the plasma membrane forming a so-called 'signalosome'. The ensuing depletion of the cytoplasmic pool of AXIN1 inhibits the formation of the destruction complex. β-catenin thereby stabilizes and translocates to the nucleus, where it interacts with TCF/LEF transcription factors activating transcription of specific target genes, including the negative feedback regulator *AXIN2*.

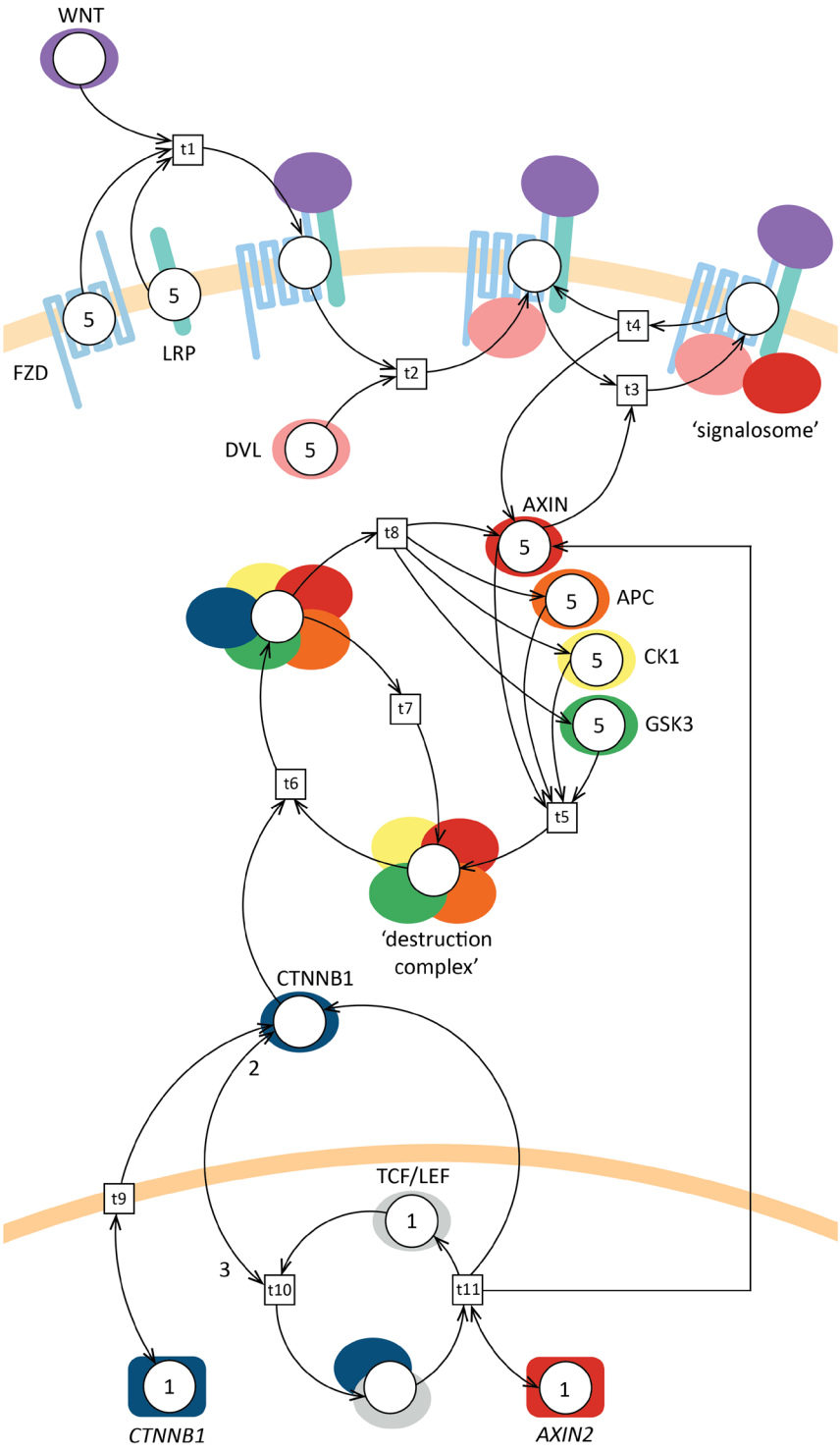
for ubiquitination (Aberle et al., 1997) and degradation (Figure 2.1A). Extracellular WNT binds to and activates the 7 transmembrane receptor, Frizzled (FZD) (Bhanot et al., 1996), and the co-receptor, lipoprotein receptor-related protein (LRP5/6) (Tamai et al., 2000). The intracellular tail of FZD interacts with Dishevelled (DVL) through an incompletely understood mechanism and sequesters AXIN1 to the cell membrane (Schwarz-Romond et al., 2007) forming a so-called 'signalosome' (Bilic et al., 2007). This leads to depletion of the cytoplasmic pool of the destruction complex component AXIN1, which in turn inhibits the formation of the destruction complex itself (Figure 2.1B). It is not fully understood whether only AXIN1 or more destruction complex components are sequestered to the cell membrane during WNT signaling. Indeed, a study by Li et al. (Li et al., 2012) showed that AXIN1 does not dissociate from the other destruction complex components during WNT signaling.

The inhibition of the destruction complex leads to β-catenin stabilization and nuclear translocation. Nuclear β-catenin interacts with TCF/LEF transcription factors (Behrens et al., 1996) forming the β-catenin/TCF transcriptional (co) activator complex. A collection of more than 100 genes induced by β-catenin/TCF transcription is listed on the WNT homepage (Nusse). The specific subset of genes induced, however, strongly depends on tissue type and developmental stage (Buchert et al., 2010). Several of these target genes are feedback regulators, where AXIN2 is of particular interest. First, *AXIN2* is a universal β-catenin/TCF target gene and as such it is believed to faithfully report Wnt-pathway activity in multiple tissues (Lustig et al., 2002; Jho et al., 2002). Second, AXIN2 encodes a functional homolog of the destruction complex component AXIN1 (Behrens et al., 1998) and mediates an auto-inhibitory feedback loop. Although AXIN1 and AXIN2 share functional similarities, they are only partially redundant *in vivo* due to their different expression patterns (Chia and Costantini, 2005): *AXIN1* is constitutively expressed (Zeng et al., 1997), whereas *AXIN2* is induced during active Wnt/β-catenin signaling (Jho et al., 2002; Leung et al., 2002). The AXIN2 negative feedback is believed to be important for the tight spatio-temporal regulation of Wnt/β-catenin signaling (Aulehla and Herrmann, 2004). However, the exact regulatory role of AXIN2 remains an open question.

Deregulated Wnt/β-catenin signaling caused by genetic alterations can have major developmental consequences, and is the leading cause of colorectal carcinogenesis (MacDonald et al., 2009). The most common colorectal cancer mutation is found in APC (Polakis, 2012; Anastas and Moon, 2013). Different APC inactivating mutations lead to different levels of Wnt-pathway activity e.g. higher β-catenin stabilization is seen in colorectal cancer compared to breast cancer (as reviewed in (Albuquerque et al., 2011)). Other rarer colorectal cancer mutations (Fearon, 2011) are found in AXIN1 (Satoh et al., 2000), AXIN2 (Liu et al., 2000; Mazzoni and Fearon, 2014) and β-catenin (Morin et al., 1997; Rubinfeld et al., 1997). As a common mode of action, these oncogenic mutations cause hyperactive WNT signaling (Klaus and Birchmeier, 2008).

Investigating the mechanism and effect of β-catenin stabilization during physiological (active) and pathophysiological (hyperactive) WNT signaling is crucial for developing effective treatment, both in the field of cancer research and regenerative medicine. *In vitro* experiments in which cells are stimulated with WNT are generally assumed to represent active signaling, whereas downstream oncogenic mutations represent hyperactive signaling. Inhibition of GSK3 using small molecule inhibitors is widely used to activate WNT signaling during cellular reprogramming and in embryonic stem cell cultures (Ying et al., 2008; Li et al., 2011). Inhibition of GSK3 inhibits the destruction complex, which can be interpreted as similar to the effects of oncogenic mutations. Several quantitative mathematical models of Wnt/β-catenin signaling have been created as reviewed in (Lloyd-Lewis et al., 2013; Kofahl and Wolf, 2010) to facilitate these investigations. The first model (Lee et al., 2003) described the cytosolic interactions in WNT signaling based on data from experiments using *Xenopus* extracts. Later models included amongst other, extensions with *i*) AXIN2 feedback, explaining effects of mutations in colorectal cancer (Cho et al., 2006); *ii*) AXIN2 feedback and another negative upstream feedback, demonstrating how these feedbacks ensure robust oscillations (Wawra et al., 2007); *iii*) WNT inhibitors, secreted Frizzled-related protein and Dickkopf, which showed a synergistic effect on β-catenin accumulation (Kogan et al., 2012); *iv*) effects of different target gene regulations induced by different WNT and APC concentrations (Benary et al., 2013); *v*) the interplay between adhesive and transcriptional functions of β-catenin (van Leeuwen et al., 2007) and *vi*) data from experiments using mammalian cells, showing significant differences in AXIN levels (Tan et al., 2012). However, constructing such quantitative models remains a challenge, because they require detailed information on e.g. protein concentrations and reaction rates. In addition to being dependent on large experimental efforts, these data are particularly difficult to obtain for a signaling pathway that does not involve a typical kinase cascade. Consequently, these models include many estimated parameters, which limits their scale of applicability (Lloyd-Lewis et al., 2013). Petri net models, on the other hand, use coarse-grained data describing currently known interactions and relative protein levels (Bonzanni et al., 2013; Bonzanni et al., 2009). A clear advantage of this is that such coarse-grained data are more readily available, and much easier to obtain. Coarse-grained Petri net models thus expand the scale of applicability for future modeling purposes, including extensions to novel signaling components or pathways.

Here we have used a combined computational and experimental approach to build the first Petri net model of Wnt/β-catenin signaling. Compared to the aforementioned mathematical models, it includes a larger number of core Wnt-pathway components. Most importantly, we have included a signaling module that captures the interactions occurring at the cell membrane. As a result, our model allows both physiological (i.e. WNT ligand induced) and pathophysiological (i.e. induced by oncogenic mutations in downstream signaling components) to be captured. Furthermore, our model offers the possibility of including the negative



(Figure 2.2. Caption on the following page).

Figure 2.2. Petri net model of Wnt/β-catenin signaling. The model consists of 18 places (circles, representing gene or protein states), 11 transitions (boxes, representing protein complex formation, dissociation, translocation or gene expression) and 41 arcs (arrows, representing the direction of flow of the tokens). WNT initiates signaling by binding to FZD and LRP (t1), forming the WNT/FZD/LRP complex. DVL and AXIN1 then interact with this complex intracellularly (t2 and t3, respectively) forming a so-called 'signalosome'. The signalosome dissociates (with a rate of once every 10 steps) into WNT/FZD/LRP/DVL and AXIN1 (t4). Note that β-catenin is referred to by its official gene name CTNNB1 in the figure. The β-catenin protein is produced every step (t9) by the *CTNNB1* gene. AXIN1, APC, CK1 and GSK3 interact (t5) and form a 'destruction complex'. The destruction complex binds β-catenin (t6) to mark it for degradation. The destruction complex is then either reused (t7) for another round of β-catenin degradation or dissociates (t8) into its components AXIN1, APC, CK1 and GSK3. Alternatively, β-catenin may interact with TCF/LEF in the nucleus (t10), leading to transcriptional activation of *AXIN2* (t11). Initial token levels are 0 (not shown), 1 or 5 (depicted in the places). Most arc weights are 1 (not shown), except for the nuclear translocation and interaction of β-catenin to TCF/LEF transcription factors, which has an incoming arc weight of 3 and an outgoing arc weight of 2 (depicted on the arcs).

feedback by AXIN2. We used the model to explain how active signaling upon WNT stimulation and hyperactive signaling upon GSK3 inhibition leads to different levels of β-catenin stabilization. We corroborated our observations from the model using data from TCF/LEF luciferase reporter assays and Western blot analysis. We then used the experimentally validated model to explore plausible modes of action of β-catenin stabilization as a result of negative feedback by activating expression of AXIN2 upon WNT stimulation, or due to APC inactivating mutations that are known to play a key role in oncogenesis of colorectal and breast cancer.

2.2 Results

2.2.1 Building a Petri net model for Wnt/β-catenin signaling

A Petri net model can be graphically represented with two types of nodes: 'places', describing the biological components, and 'transitions', describing the activity between the biological components, which can be constructed based on known interaction data. The places are denoted with 'tokens', which describe the relative availability of the biological component. Tokens are assigned on the basis of existing data from the literature, typically relative protein levels. Places and transition are connected by weighted arcs that are important for the flow of the network. (Krepska et al., 2008; Petri, 1962; Reisig and Rozenberg, 1998) (see Materials and Methods for a detailed explanation of Petri nets).

We created a Petri net model of Wnt/β-catenin signaling to investigate the mechanism and effect of β-catenin stabilization under physiological (e.g. embryonic development) and pathophysiological (e.g. cancer) conditions based on the literature as described in details below (Figure 2.2). The model describes the interactions between the core proteins in the pathway with a focus on capturing the immediate behavior of β-catenin stabilization following destruction complex inactivation. Therefore the degradation of β-catenin by the destruction complex is specifically

included in the model, while no changes in the production/degradation rate of the other proteins needed to be assumed. Likewise, the gene expressions of *β-catenin* (encoding β-catenin), but also of *AXIN2*, a negative feedback target gene of the Wnt-pathway, are also specifically included in the model. Production and degradation of all other proteins are assumed to have similar rates and are therefore omitted, such that the token levels of these proteins remain the same throughout the simulation (as detailed in Materials and Methods).

The final model consists of 18 places (circles, representing gene or protein states), 11 transitions (boxes, representing protein complex formation, dissociation, translocation or gene expression) and 41 arcs (arrows, representing the direction of flow of the tokens). In the model, WNT initiates signaling extracellularly by binding to its transmembrane receptors FZD and LRP (t1), forming the WNT/FZD/LRP complex. DVL interacts with the intracellular tail of FZD when present in the WNT/FZD/LRP complex (t2), forming the WNT/FZD/LRP/DVL complex. DVL thereafter sequesters AXIN1 to the membrane (t3) forming the signalosome consisting of WNT, FZD, LRP, DVL and AXIN1. In the model we have not specifically included the contribution of GSK3 and CK1 in the formation of the signalosome, because these two multi-tasking kinases are generally assumed not to be rate-limiting in the cell (MacDonald et al., 2009; Lee et al., 2003). Further, AXIN1 is the only destruction complex constituent that binds to the signalosome in the model. The signalosome dissociates once every 10 steps (t4) into the WNT/FZD/LRP/DVL complex and AXIN1 in order to incorporate a lower dissociation- than formation-rate of the signalosome. The destruction complex, which sequesters β-catenin unless WNT induces signalosome formation, is formed (t5) by AXIN1, APC, CK1 and GSK3. In the model, β-catenin binding to the destruction complex leads to degradation of β-catenin (t8 and t7), and the destruction complex is then either reused (t7) for another round of β-catenin degradation or dissociates (t8) to AXIN1, APC, CK1 and GSK3. In the model, β-catenin protein is produced every step (t9) following transcription of the *β-catenin* gene, and either binds the destruction complex (t6) or translocates to the nucleus, where it interacts with TCF/LEF (t10) to activate transcription of *AXIN2* (t11). Since AXIN1 and AXIN2 are functional homologs (Behrens et al., 1998), they are modeled as one protein entity (depicted as 'AXIN'). Further, we do not distinguish between the cytoplasmic and nuclear pool of β-catenin in the model. This allowed the nuclear translocation and TCF/LEF interactions to be modeled as one transition (t10).

The initial token level of the protein places were set to 5 (with exceptions described below) in order to have a concise but sufficiently large pool of the protein tokens for the model to work with. Higher levels had no effect, and lower levels restricted the expressiveness of the model, i.e. too coarse-grained. The initial token level of the β-catenin protein place and all protein complexes (WNT/FZD/LRP, WNT/FZD/LRP/DVL, signalosome, destruction complex-β-catenin and TCF/LEF-β-catenin) were set to 0, because these are produced throughout the simulation and are assumed to be initially absent. The initial token level of TCF/LEF was set

to 1 which gives the lowest rate of TCF/LEF-β-catenin interaction, and hence the induction of low levels of AXIN2 (de la Roche et al., 2008). A parameter sweep of initial token levels of TCF/LEF confirmed that higher initial token levels resulted in a stronger response and higher AXIN2 levels. Initial token levels of the gene places, *β-catenin* and *AXIN2*, were set to 1 and kept at these values since these genes are always presumed to be present. In our model this means that *AXIN2* is induced only when the TCF/LEF-β-catenin complex is present and *β-catenin* is induced once every step, because there is nothing implemented to restrain this induction (β-catenin levels are instead regulated by the destruction complex).

Most arc weights were set to 1, which means that the model dynamics rely on its connectivity. This has been proven successful using Petri net modeling of MAPK and AKT signaling cascades (Bonzanni et al., 2013; Bonzanni et al., 2009; Ruths et al., 2008). An exception to this was applied when implementing fractional arc weights to represent a lower transition rate (i.e. the dissociation of the signalosome, the effects of different APC mutations and the induction of AXIN2) and when implementing a higher interaction affinity of β-catenin to the destruction complex than to TCF/LEF. For the latter the arc weight from β-catenin to t10 (i.e. its translocation to the nucleus and subsequent interaction with TCF/LEF) was set to 3, and the arc weight from t10 to β-catenin was set to 2. From the model point of view this means that for t10 to fire, the β-catenin place needs to be occupied by 3 tokens, but only 1 is consumed (See Figure 2.2). These weights were chosen because it is generally assumed that β-catenin accumulates in the cytoplasm before it translocates to the nucleus and binds TCF/LEF (as reviewed in (Jamieson et al., 2014)). This assumption is based on the higher interaction affinities of β-catenin to the destruction complex compared to TCF/LEF (as reviewed in (Harris and Peifer, 2005)), which means that β-catenin will interact with TCF/LEF when the β-catenin levels begin to increase because the destruction complex is unable to keep up with the degradation. In order to implement the lower dissociation rate of the signalosome, the ingoing and outgoing arcs of t4 were implemented with a fractional arc weight of 0.1, which represents a firing rate of once every 10 steps. Further, we implemented a constraint on the transition to ensure that it does not fire more than once every step. Parts of this initial setup were changed accordingly to mimic the different conditions of Wnt/β-catenin signaling simulated in this study (see below).

2.2.2 Active signaling upon WNT stimulation

We simulated WNT stimulation to predict the level of β-catenin stabilization during active signaling. To this end we ran a series of simulations with different initial WNT token levels (ranging from 0 to 5) without AXIN2 feedback (i.e. the arc weight from t11 to AXIN was set to 0). As shown in Figure 2.3A, we observed four different β-catenin response levels depending on the initial WNT token level. A flat β-catenin response was seen for WNT = 0, 1 or 2. For WNT = 3, 4 or 5, we observed a delay in the initial increase of β-catenin, which eventually increased linearly with a slope

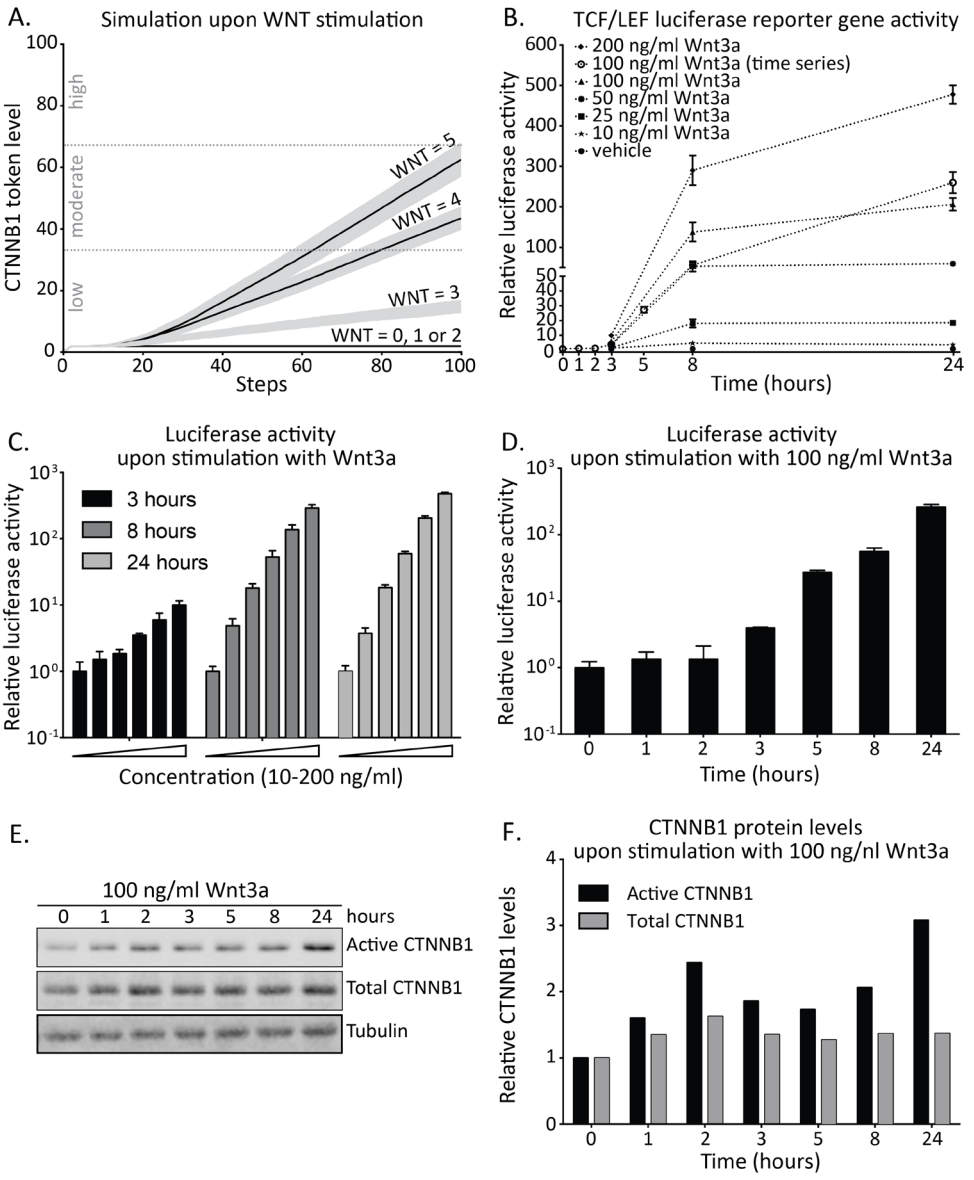


Figure 2.3. Model simulation and experimental validation of Wnt-pathway activation upon WNT stimulation. (A) β-catenin (referred to by its official gene name CTNNB1 in the figure) token levels predicted by our model with initial WNT token levels ranging from 0 to 5. For WNT = 0, 1 or 2, we observed a flat β-catenin response. For WNT = 3, 4 and 5 β-catenin increases from low to moderate levels. (B) Graph combining the results from panels C and D to allow easy comparison to the modeling results depicted in panel (A), showing dose- and time-dependent activation of a Wnt/β-catenin responsive TCF/LEF luciferase reporter in HEK293T^{WOO} cells. For all curves with black data points (corresponding to panel C), luciferase activity was plotted relative to the vehicle control (not shown), which was set at 1 for each of the three time points (3, 8 and 24 hours). For the curve with white data points (corresponding to panel D), luciferase activity was plotted relative to the vehicle

(Continuing caption for Figure 2.3 on previous page).

control, which was set at 1 for the t=0 hours condition. (C) Reporter assay in HEK293T^{WOO} cells, showing dose-dependent activation at 3, 8 and 24 hours after stimulation with purified Wnt3a (same concentrations as depicted in B). (D) Reporter assay in HEK293T^{WOO} cells, showing time-dependent activation upon treatment with 100 ng/ml of Wnt3a. Values were plotted relative to the vehicle control, which was set at 1 for t=0 hours. (E) Western blot from the experiment depicted in (D), showing total and active (i.e. non-phosphorylated) β-catenin levels. Since the soluble, signaling pool of β-catenin constitutes only a minor fraction of the total pool of β-catenin, the use of antibody against active β-catenin ensures that only the pool involved in WNT signaling is visualized. Tubulin was used as a loading control. (F) Quantification of the Western blot shown in (E). Total and active β-catenin levels were normalized to tubulin. The increase in either total or active β-catenin levels was plotted relative to time point 0, for which the normalized levels were set to 1. Experiments were repeated two (C) or three (D-F) times. A representative experiment is shown.

depending on the WNT level. The β-catenin stabilization was low for WNT = 3 and moderate for WNT = 4 and 5. Maximal WNT stimulation (WNT = 5) led to a stabilization of ~60 β-catenin tokens.

To compare the response predicted by our model to the biological response of cells treated with Wnt3a, we measured β-catenin activation by both Western blot analysis and TCF/LEF luciferase reporter assay (Figure 2.3B-2.3F and Figure S2.1). Although the former directly detects β-catenin levels, the latter faithfully reports Wnt/β-catenin signaling (Molenaar et al., 1996) and remains the most sensitive and robust method to quantify Wnt/β-catenin signaling to date (Nusse; Veeman et al., 2003; van de Wetering et al., 1997). Furthermore, it allows high-throughput analyses of Wnt-pathway activation (i.e. a comparison of multiple doses and time points within the same experiment). To validate the β-catenin levels predicted upon WNT stimulation by our model, we treated HEK293T^{WOO} cells (carrying a stably integrated β-catenin/TCF luciferase reporter) with increasing concentrations of purified, commercially available, Wnt3a for 3, 8 and 24 hours. These experiments reproduce the dose- and time-dependent increase of TCF/LEF reporter gene activity predicted by our model (Figure 2.3B and 2.3C). To directly link the results from the reporter gene assay to an increase in β-catenin protein levels, we repeated the experiment for one level of WNT stimulation (100 ng/ml purified Wnt3a) for a more extensive time series, including additional earlier time points, and analyzed the results by performing both a TCF/LEF reporter gene assay (Figure 2.3B and 2.3D) and quantitative Western blot analysis (Figure 2.3E and 2.3F), which allows direct, albeit less sensitive, detection of β-catenin protein levels. Both the transcriptional reporter assay and the measurement of β-catenin protein levels show a time-dependent increase (Figure 2.3D-2.3F). Direct comparison of the two readouts reveals the inherent limitations of each of the two experimental systems: The change (i.e. fold increase) in TCF/LEF reporter activity is more pronounced than, but slightly delayed compared to, the change in β-catenin protein levels. Our Petri net model (Figure 2.3A) shows the same qualitative effect: a consistent rise in β-catenin levels. As such, our model more closely mimics the luciferase response (i.e. activation of an artificial reporter gene).

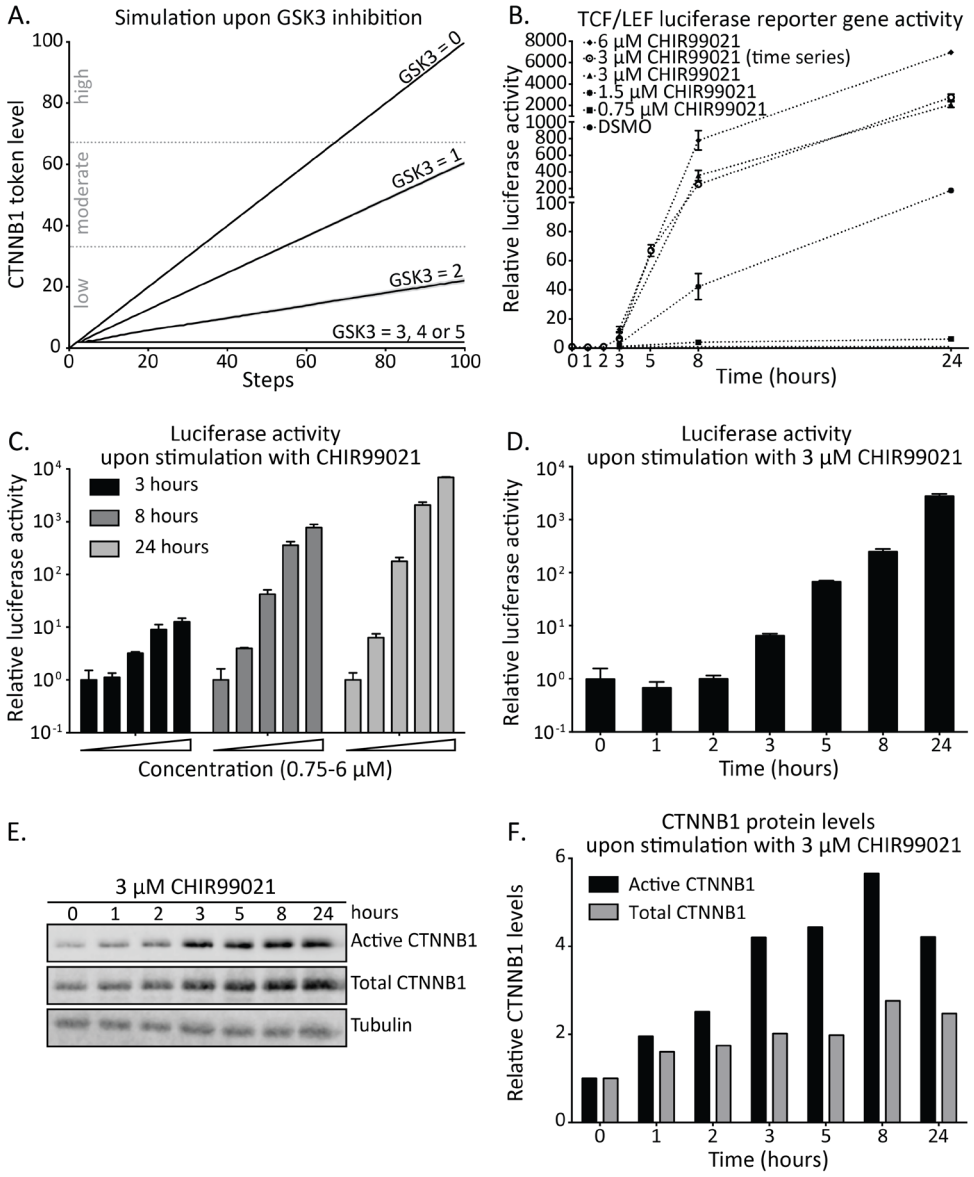


Figure 2.4. Model simulation and experimental validation of Wnt-pathway activation upon GSK3 inhibition. (A) β-catenin (referred to by its official gene name CTNNB1 in the figure) token levels predicted by our model with initial GSK3 token levels ranging from 0 to 5. For GSK3 = 3, 4 or 5, we observed a flat β-catenin response. For GSK3 = 0, 1 or 2 β-catenin increases to low, moderate or high levels, respectively. (B) Graph combining the results from panels C and D to allow easy comparison to the modeling results depicted in panel (A), showing dose- and time-dependent activation of a Wnt/β-catenin responsive TCF/LEF luciferase reporter in HEK293T^{WOO} cells. For all curves with black data points (corresponding to panel C), luciferase activity was plotted relative to the vehicle control (not shown), which was set at 1 for each of the three time points (3, 8 and 24 hours). For the curve with white data points (corresponding to panel D), luciferase activity was plotted relative to the

(Continuing caption for Figure 2.4 on previous page).

vehicle control, which was set at 1 for the t=0 hours condition. (C) Reporter assay in HEK293T^{WOO} cells, showing dose-dependent activation at 3, 8 and 24 hours after stimulation with CHIR99021 (same concentrations as depicted in B). (D) Reporter assay in HEK293T^{WOO} cells, showing time-dependent activation upon treatment with 3 µM CHIR99021. Values were plotted relative to the DMSO control, which was set at 1 for t=0 hours. (E) Western blot from the experiment depicted in (D), showing total and active (non-phosphorylated) β-catenin levels. Tubulin was used as a loading control. (F) Quantification of the Western blot shown in (E). Total and active β-catenin levels were normalized to tubulin. The increase in either total or active β-catenin levels was plotted relative to time point 0, for which the normalized levels were set to 1. Experiments were repeated two (C) or three (D-F) times. A representative experiment is shown.

2.2.3 Hyperactive signaling upon GSK3 inhibition

To predict the level of β-catenin stabilization during hyperactive signaling by a downstream perturbation, we next simulated our model upon GSK3 inhibition. We ran a series of simulations with different initial GSK3 token levels (ranging from 5 to 0), where 5 initial tokens represents wildtype (i.e. no Wnt-pathway activity) and 0 corresponds to complete inhibition (hyperactive signaling). The simulations revealed that the response levels depend on initial GSK3 token levels (see Figure 2.4A). For GSK3 = 3, 4 or 5, we observed a flat β-catenin response. A linear increase in β-catenin levels with a slope depending on GSK3 levels was seen for GSK3 = 0, 1 or 2. This corresponds to β-catenin degradation ranging from no degradation to 1 or 2 β-catenin tokens degraded per three simulation steps, respectively. Consequently, β-catenin stabilization was low for GSK3 = 2, moderate for GSK3 = 1 and high for GSK3 = 0. Complete GSK3 inhibition led to a stabilization of 100 β-catenin tokens.

To validate the coarse-grained β-catenin levels predicted by our model upon GSK3 inhibition, we stimulated HEK293T^{WOO} cells with increasing concentrations of CHIR99021, one of the most potent and selective GSK3 inhibitors available to date (Ring et al., 2003), over a broad time range (3, 8 and 24 hours). The measured TCF/LEF reporter gene activity confirmed the dose- and time-dependent increase upon GSK3 inhibition (Figure 2.4B and 2.4C) predicted by our model (Figure 2.4A). As with the Wnt3a treatment, here we also performed a TCF/LEF reporter gene assay and quantitative Western blot analysis side by side for one of the treatment conditions (3 µM CHIR99021) for multiple time points. The 3 µM CHIR99021 concentration was chosen in order to achieve a near maximal GSK3 inhibition or Wnt-pathway activation. An increase in both active (i.e. non-phosphorylated) and total (i.e. both phosphorylated and non-phosphorylated) β-catenin is apparent after 1 hour, whereas an increase in the signal of the luciferase reporter assay can only be detected after 3 hours. Furthermore, the dynamic range of the Western blot analysis is limited compared to the reporter gene assay, allowing us to measure at most a 4-fold increase in β-catenin levels in the former, but up to a 10^4 fold increase in Wnt-pathway activity in the latter (Figure 2.4D-2.4F).

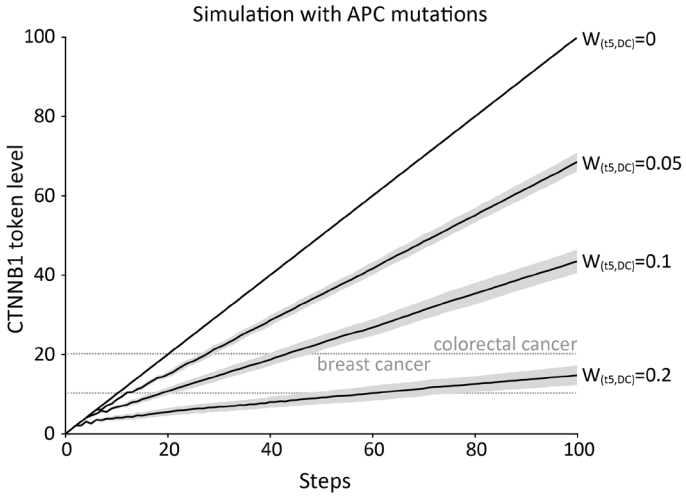


Figure 2.5. Model prediction of Wnt-pathway hyperactivation by APC inactivating mutations. β-catenin (referred to by its official gene name CTNNB1 in the figure) token levels predicted by our model with four different APC mutations (weight on the arc going from t5 to the destruction complex (DC) set at 0.2, 0.1, 0.05 and 0, respectively). Moderate and high β-catenin stabilizations might correspond to the effects by mutations in colorectal tumor formation and the low β-catenin stabilizations might correspond to the effects by mutations in breast tumor formation.

2.2.4 Predictions of hyperactive signaling by APC inactivating mutations

The most common colorectal oncogene, *APC*, perturbs downstream WNT signaling. Different APC mutations exist that result in truncated proteins negatively influence the formation of the destruction complex to different degrees. As a result, the different APC mutations lead to different levels of β-catenin stabilizations. According to a recent review (Albuquerque et al., 2011), the β-catenin signaling activity (β-catenin reporter activity) was low (between 10-20%) for APC mutations in breast tumors, versus moderate to high (between 20-100%) in colorectal tumors.

We used our validated model to explore if the effect of these APC mutations might be explained by different rates of destruction complex formation. We implemented the effect of the APC mutations by decreasing the rate of the destruction complex formation ranging from no production at all to production every 20, 10 and 5 steps. In Figure 2.5 we observed four different response levels for the different APC mutations, where stabilization of β-catenin levels went from low to high depending on this rate of destruction complex formation. Comparing these token levels to the β-catenin signaling activities reviewed in (Albuquerque et al., 2011), the three highest β-catenin stabilizations would correspond to hyperactive signaling by APC mutations in colorectal tumor formation (moderate and high β-catenin token levels), whereas the lowest β-catenin stabilization would correspond to the effects by APC mutations as observed in breast tumor formation (low β-catenin token levels).

2.2.5 Predictions of active signaling upon WNT stimulation with AXIN2 feedback

In our model *AXIN2* is induced by β-catenin/TCF transcription and increases the cytoplasmic pool of AXIN, which under certain conditions, e.g. WNT stimulation, is the limiting factor for β-catenin degradation. However, our experimental dataset obtained using Wnt3a stimulation showed no obvious decrease in β-catenin levels that might be due to this negative feedback (Figure 2.3E and 2.3F). It should be noted that in this experimental setting (100 ng/ml Wnt3a), the Wnt-pathway is likely still activated at supra-physiological levels. Moreover, the WNT ligand remains present throughout the experiment. *In vivo*, however, physiological Wnt-pathway activation is strictly regulated both due to the WNT concentration gradient and due to the tight spatio-temporal control of Wnt gene expression. Under these circumstances, lower levels of Wnt/β-catenin signaling are likely to occur and, as a result, part of the regulation may be due to the AXIN2 auto-inhibitory feedback loop. Therefore, it may be the ratio between the WNT and AXIN2 levels that is crucial to the regulatory role of AXIN2. We therefore used our model to get further insights into the effect of the AXIN2 feedback and explored a spectrum of possible β-catenin stabilizations under different WNT and AXIN2 levels. We ran a series of simulations with different initial WNT token levels: 3, 4 or 5, which showed increased β-catenin stabilization in Figure 2.3A, and with different AXIN2 feedback strengths: the arc weight from t11 to AXIN was varied from 0 for no feedback to 0.15 for maximum feedback. As shown in Figure 2.6, we observed three different spectra of β-catenin stabilizations depending on the different initial WNT token levels. The highest β-catenin stabilizations (solid lines in Figure 2.6) were identical to those observed in Figure 2.3A (without AXIN2 feedback). At high feedback, the β-catenin stabilization is lowered, and a maximum appears after which the β-catenin level declines (dashed lines in Figure 2.6). The lowest β-catenin stabilizations displayed three different peak responses. For the peak responses, the height of the peak and the duration of the response depended on initial WNT token levels. Maximal β-catenin stabilization comes later in the simulation for higher initial WNT token levels.

2.3 Discussion

In spite of more than 30 years of study, the Wnt/β-catenin signaling pathway still holds many questions. The molecular details of how an external WNT stimulus results in the stabilization of transcriptionally active β-catenin/TCF complexes remain incompletely understood and, in some cases, a topic of debate (Verkaar et al., 2012). The Petri net model presented in this paper allows us to investigate downstream effects of Wnt/β-catenin signaling under physiological and pathophysiological conditions. The core of the model describes the current state of knowledge as summarized in the introduction. The coarse-grained nature of the interaction and protein level data used to build our model allowed us to include

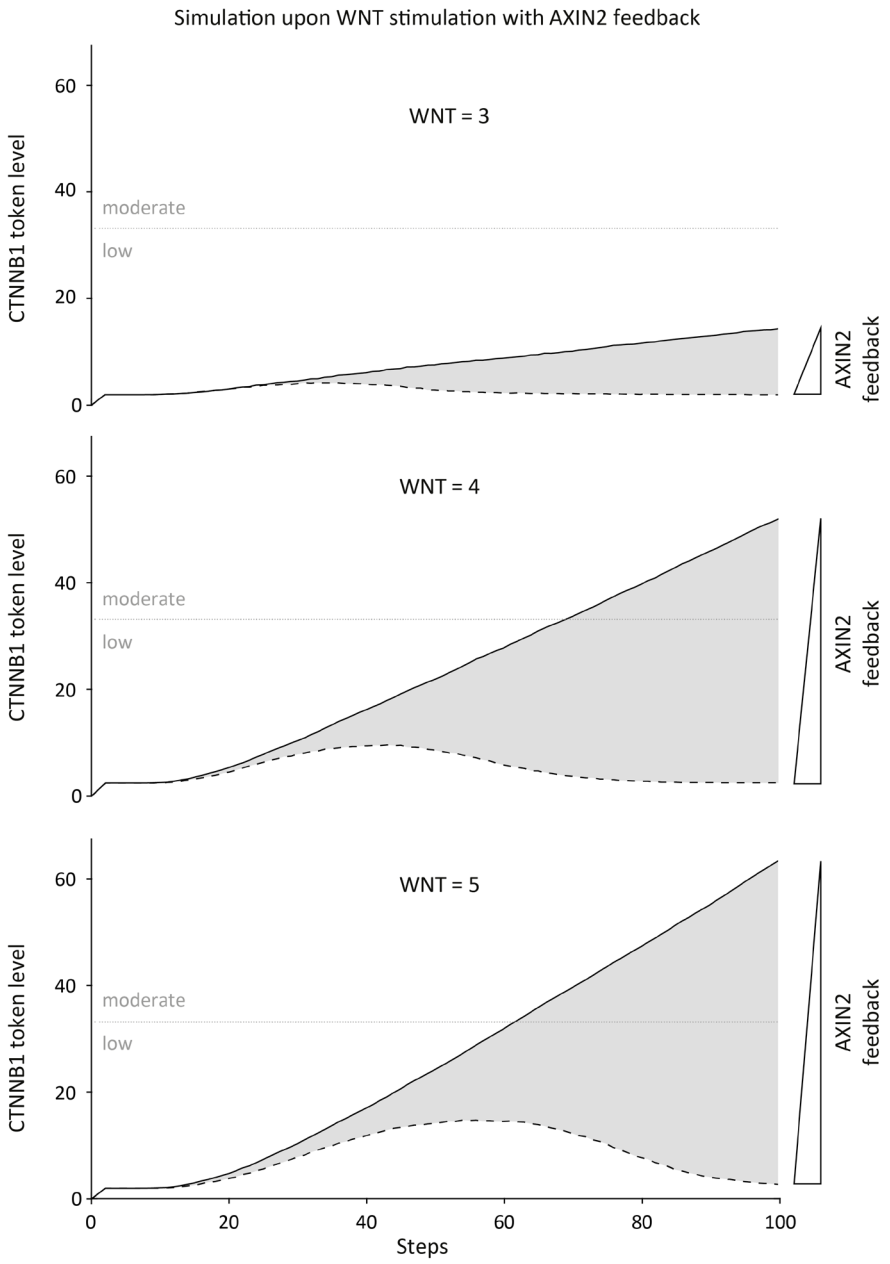


Figure 2.6 Model prediction of Wnt-pathway activation upon WNT addition with AXIN2 feedback. β-catenin (referred to by its official gene name CTNNB1 in the figure) token levels predicted by our model with arc weight from t11 to AXIN varied from 0 (no feedback; solid lines) to 0.15 (high feedback, dashed lines), and initial WNT token levels at 3, 4 and 5 (top, middle and bottom panels, respectively). We observed three spectra of β-catenin stabilizations depending on initial WNT levels. The highest β-catenin stabilizations correspond to simulations without AXIN2 feedback (solid lines), whereas with high AXIN2 feedback the β-catenin stabilization was attenuated (dashed lines).

the interactions at the plasma membrane, which is lacking from currently existing quantitative models as reviewed in (Lloyd-Lewis et al., 2013; Kofahl and Wolf, 2010). Of note, AXIN1 is the only destruction complex component sequestered to the plasma membrane during Wnt/β-catenin signaling in our model. For now we ignored the actions of GSK3 at the plasma membrane, where it phosphorylates LRP5/6 to create the AXIN1 binding site and where, in turn, the kinase itself may be inhibited (Davidson et al., 2005; Zeng et al., 2005). Furthermore, we considered free cytoplasmic and nuclear β-catenin as a single pool in the model. More detailed experimental data on subcellular compartmentalization of β-catenin (or any other signaling component) would allow us to refine our Petri net model, which easily allows incorporation of such detail. For instance, at present our model only considers the active (i.e. unphosphorylated), rather than the total pool of β-catenin, which would also include the β-catenin present in the adhesion complexes in the cell membrane.

Our Petri net model for Wnt/β-catenin signaling was constructed based on known interactions of signaling components described in the literature, thereby capturing the current state of knowledge in the field. To validate the model, we generated our own experimental data. This allowed a direct comparison of physiological (i.e. WNT stimulation) and pathophysiological (i.e. GSK3 inhibition) activation of the pathway in the same cell line, using time points and analyses best suited for connecting our experimental data and the Petri net modeling predictions.

Our model predicts a dose- and time dependent response for both WNT stimulation and GSK3 inhibition (Figures 2.3A and 2.4A). This is confirmed by the experimental data (Figures 2.3B, 2.3C, 2.4B and 2.4C). The main discrepancy between the simulated and the experimental data is the time-delay that is predicted in response to WNT stimulation compared to GSK3 inhibition (compare Figure 2.3A to 4A). Indeed, activation of Wnt/β-catenin signaling is known to be a slow event (unlike the activation of MAPK signaling for instance, which occurs within a matter of minutes) (Naik and Piwnica-Worms, 2007; Yokoyama et al., 2007; Verkaar et al., 2010). However, we did not detect this delay in β-catenin accumulation by either TCF/LEF luciferase reporter assay (compare Figure 2.3B-2.3D to Figure 2.4B-2.4D) or Western blot analysis (compare Figure 2.3E and 2.3F to Figure 2.4E and 2.4F). We believe this is mainly due to experimental limitations. Given that a subtle increase in β-catenin protein levels can be detected approximately one hour after stimulation with either Wnt3a (Figure 2.3E and 2.3F) or CHIR99021 (Figure 2.4E and 2.4F), any delay in activation of the Wnt-pathway must occur prior to that time point. Detecting this delay would require assays with superior spatio-temporal resolution. The delay predicted by our model upon WNT stimulation (Figure 2.3A) can be explained by the fact that formation of the signalosome occurs a few steps into the simulation, whereas inhibition of GSK3 is a one-step event, and that the transitions for signalosome formation (i.e. pathway activation) and destruction complex formation (i.e. pathway inhibition) compete for AXIN1. Thus, when AXIN1 is sequestered to the plasma membrane, less cytoplasmic AXIN1 is available

for formation of the destruction complex. To what extent these events contribute to Wnt-pathway activation under experimental conditions remains unknown, owing to the absence of tools to study the exchange of endogenous AXIN1 between these two pools. Our results do suggest that competition over AXIN1 between the destruction complex and the signalosome may well be important also under physiological conditions.

The time-delay together with the continuous sequestration and dissociation of AXIN1 to the signalosome leads to prediction of higher stabilization of β-catenin for complete GSK3 inhibition compared to maximal WNT stimulation, where the difference is almost two-fold (compare Figure 2.4A to Figure 2.3A). We observe a similar difference when measuring TCF/LEF reporter gene activity: the highest concentration of CHIR99021 activates the reporter approximately 10-fold higher than the highest concentration of Wnt3a tested (compare Figure 2.4B-2.4D to Figure 2.3B-2.3D). Comparing protein levels, instead of transcriptional activation, shows a much smaller difference: 2-fold higher β-catenin at most when cells are stimulated with CHIR99021 versus Wnt3a (compare Figure 2.4E and 2.4F to Figure 2.3E and 2.3F). Although it is tempting to conclude that this data again confirms the predictions of our model, it should be stressed that the different experimental modes of Wnt-pathway activation cannot be compared directly. This is because they are achieved by different molecules (i.e. purified Wnt3a versus a synthetic small-molecule GSK3 inhibitor) with different intrinsic activities and chemical properties such as half-life and stability in the tissue culture medium, which may greatly impact on the experimental outcome. At the same time, we may speculate that the observed differences reflect real differences in sensitivity of the Wnt-pathway. In this case, our experimental findings might be explained by the fact that the more physiological means of pathway activation by Wnt3a is more likely to be subject to negative feedback control via AXIN2 induction than the more artificial perturbation by CHIR99021 inhibition of GSK3 at the level of the destruction complex.

AXIN2 is one of the few comprehensive globally expressed WNT target genes and is thought to act as a negative regulator to Wnt/β-catenin signaling (Lustig et al., 2002; Jho et al., 2002). The degree to which AXIN2 attenuates WNT signaling and the actual spatio-temporal regulatory role of AXIN2 is still a topic of debate. Indeed, when we incorporate this negative feedback loop in the model upon WNT stimulation, our simulations predict that Wnt-pathway activity is attenuated (at certain levels of AXIN2 induction), and ultimately returns to baseline levels (dashed lines in Figure 2.6). Importantly, in the model the feedback from AXIN2 only negatively influences stabilized β-catenin levels when AXIN1 is the limiting factor. This is the case when AXIN1 is deprived from the cytoplasm by sequestration to the signalosome (i.e. upon WNT stimulation). This is why we are able to observe a negative effect from the AXIN2 feedback upon WNT stimulation in our model (Figure 2.6), but not by GSK3 inhibition (Figure 2.4A) or APC inactivating mutations (Figure 2.5). It should be noted however, that on the

timescale used for the experiments, we do not observe complete feedback inhibition by AXIN2 (Figure 2.3B-2.3F). This might be due to the relatively low level of AXIN2 induction in the cells used for these experiments (de la Roche et al., 2008) in combination with supra-physiological levels of Wnt-pathway activation achieved upon stimulation with purified Wnt3a. However, it could also be due to the fact that AXIN1 is not the limiting factor in the cells used for this study. Previously, a study of Wnt/β-catenin signaling in *Xenopus laevis* showed that AXIN1 is 1000-fold lower than the other components of the destruction complex (Lee et al., 2003) and has therefore been considered the natural limiting factor. However, a recent study of Wnt/β-catenin signaling in mammalian cells showed that the concentrations of the components of the destruction complex were on the same range (Tan et al., 2012). Therefore, we cannot exclude the possibility that AXIN1 is not the limiting factor in the cells used for this study. Unfortunately, the current experimental tools, most notably Western blot analysis of endogenous β-catenin levels, are not robust, high-throughput and sensitive enough to resolve this issue. However, by using our model we were able to predict and visualize spectra of β-catenin stabilization, which showed that the ratio between the WNT and AXIN2 levels are important for the degree of feedback observed (Figure 2.6). The two most notable observations were that, for high WNT levels, a higher level of AXIN2 was needed to reach baseline β-catenin levels and, for low WNT levels, a baseline β-catenin level is reached early. Based on these predictions we can speculate whether the AXIN2 negative feedback only has an effect on low WNT levels and whether the regulatory role of this is to insure a faster on/off switch of Wnt-pathway activity. At present, in both our simulations and current experimental setting the cells are continuously exposed to Wnt3a. During normal development cells may essentially only receive a short pulse of WNT stimulation, given that the hydrophobic WNT proteins either do not travel far from their production source or are quickly sequestered by responsive cells. Indeed, in vivo Wnt-pathway activity shows dynamic on and off switches during development (Aulehla and Herrmann, 2004; Gonzalez et al., 2013; Suriben et al., 2006). Examples of these are the restriction of Wnt/β-catenin responsive cells to the crypt, but not to the villus sections of the intestinal epithelium, and oscillation of WNT signaling as part of the mouse segmentation clock. By including protein degradation and different sources of WNT protein in our model, such oscillations might also be simulated.

In conclusion, our Petri net model of Wnt/β-catenin signaling provides insight on the mechanisms leading to different levels of β-catenin stabilization upon WNT stimulation and GSK3 inhibition corroborated by TCF/LEF luciferase assay and Western blot analysis. It should be stressed that the simulations show a coarse-grained output per step and we cannot directly map token levels to the relative activities in the TCF/LEF luciferase reporter assay nor to the β-catenin levels measured by Western blot analysis. Furthermore, we also cannot directly map a simulation step in the model to an experimental timescale. Despite these limitations, our model resembles Wnt/β-catenin signaling to the extent that it captures the logic

of the interactions and reflects the sequence of events of pathway activation and repression by various mechanisms. In this way, our model can be used to simulate and predict both physiological and pathophysiological WNT signaling. Thus, this modeling exercise has allowed us to study the mechanisms and effects of Wnt/β-catenin signaling under different conditions, as well as the effects of protein- and pathway-modifications that are known to influence this pathway in many types of cancer.

2.4 Materials and methods

2.4.1 Petri net modeling

We built a Petri net model of Wnt/β-catenin signaling describing known components, actions and interactions, well established in literature, in a logical way. A Petri net consists of two types of nodes, 'places' and 'transitions', and is connected by directed edges called 'arcs'. A place represents an entity (e.g. gene or protein), whereas a transition indicates the activity occurring between the places (e.g. gene expression or complex formation). Places can only link to transitions and vice versa (i.e., a Petri net is a bipartite graph). The direction of the arcs is important for the flow of the network. An arc goes from an input place to a transition, and from a transition to an output place. Places contain 'tokens', indicating the availability of the corresponding entity, while arcs have a weight, denoting the amount of tokens to consume from an input place or to produce to an output place. If the token levels of all input places of a transition fulfill the requirement of (i.e. are equal to or higher than) the weights of the respective arcs, the transition is enabled. Only enabled transitions can be executed, leading to transfer (consumption/production) of tokens between places. Note that if two (or more) enabled transitions share an input place, they may be in competition if available token levels do not allow simultaneous execution of both (or all). In our model, AXIN, β-catenin and the destruction complex with β-catenin bound, are each input places for two transitions (t3/t5, t6/t10 and t7/t8, respectively). The initial token levels and the arc weights are generally restricted to be integer values. However, to represent a lower firing rate of a transition (i.e. a rate-limiting step) a fractional arc weight is implemented.

Gene expression is modeled such that one arc goes from the gene-place to the transcriptional-transition, one arc goes from the transcriptional-transition to the gene-place, and one arc goes from the transcriptional-transition to the protein-place. Gene-places are initiated with 1 token and when the transcriptional-transition of a gene is enabled the token is produced both in the protein-place and in the gene-place itself. This way the token can be reused for another round of gene expression, reflecting the fact that the gene (DNA) is needed, but is not consumed during expression.

2.4.2 Active and hyperactive conditions in the model

We modeled active and hyperactive signaling upon WNT stimulation and GSK3 inhibition, respectively, and used these conditions to validate the model with experimental data (see below). Inhibition of GSK3 inhibits formation of the destruction complex, which we interpret to be similar to oncogenic perturbations. Therefore, for modeling purposes, GSK3 inhibition was used to mimic hyperactive signaling. For GSK3 inhibition we varied the initial token level of GSK3, respectively, from 0 to 5. For WNT stimulation we varied the initial token level of WNT from 0 to 5 and removed the AXIN2 feedback (the arc weight from t11 to AXIN was set to 0). The experimentally validated model was used to predict the level of β-catenin stabilization with the AXIN2 negative feedback upon WNT stimulation and APC inactivating mutations, respectively. Upon WNT stimulation with the AXIN2 feedback we varied the initial token level of WNT (3, 4 and 5) and the arc weight from t11 to AXIN (0 (no feedback) and 0.15 (maximal feedback)). The maximal feedback of 0.15 was chosen based on the criteria that it should i) show a peak response and ii) return to basal level for all initial WNT token levels (3, 4 and 5). This arc weight represents a firing rate of three times every 20 steps. Thus, the simulation for each initial WNT token level produced two β-catenin stabilization curves (i.e. no feedback and maximal feedback). The area between these two curves was used to explain the spectra of β-catenin stabilizations at intermediate levels of *AXIN2* induction. APC mutants have decreased binding affinity to the other components of the destruction complex to different degrees. We implemented this by reducing the formation of the destruction complex i.e. the weight on the arc going from the complex-formation-transition (t5) to the destruction complex (production) was decreased to 0, 0.05, 0.1 and 0.2. In addition, we incorporated arcs going from the complex-formation-transition to the individual destruction complex components (i.e. AXIN1, APC, GSK3 and CK1) with arc weights of 1 minus the production-weight to equally decrease the consumption. The implemented arc weight of 0 represents no production of the destruction complex (i.e. a complete null mutation). The implemented fractional arc weights of 0.05, 0.1 and 0.2 represent a production of the destruction complex once every 20, 10 and 5 steps, respectively.

2.4.3 Simulations

The model was simulated with maximally parallel execution, cf. our previous work (Bonzanni et al., 2009), where the maximum possible number of enabled transitions are executed at each simulation step. This mimics the behavior in the cell, where typically many interactions happen at the same time. Two or more transitions can compete over one input place, as mentioned above. If this place only contains enough tokens to enable one of the transitions, but not both, a conflict occurs which is resolved by randomly drawing one of the competing transitions to execute. This makes the simulations non-deterministic.

For each condition we simulated the total β-catenin token levels, i.e. free β-catenin and β-catenin bound to the destruction complex or TCF/LEF, respectively, over 100 steps repeated 100 times. To account for variations in token levels due to

the non-deterministic nature of the model, the mean and standard deviation of the β-catenin token levels over the 100 simulations were calculated for each step. These 100 steps represent the (early) time scale of β-catenin accumulation and the final token count represents the stabilized β-catenin level measured in experiments. The simulation steps describe the sequence of events and should not be linearly translated to time units. Similarly, the token level is a coarse-grained quantitative representation of actual protein levels and should not be linearly translated to a concentration. Instead, for analysis of the simulations we observe relative differences of β-catenin token levels over steps between simulations (i.e. different conditions and dosages). To validate the model we compared the relative β-catenin token levels predicted by our model simulations to the relative Wnt-pathway activities measured in the experiment (see below), where we distinguish between 'low', 'moderate' and 'high' levels. A Python script was written to run the simulations and is available together with the model in pnml format via http://www.ibi.vu.nl/downloads/WNTmodel/.

2.4.4 Cell lines

HEK293T^{WOO} (WNT OFF/ON) cells were generated by transfecting HEK293T cells with a 7×Tcf-FFluc//SV40-Puro^{R} (7TFP) reporter plasmid (a gift from Christophe Fuerer, (Fuerer and R., 2010)). Following puromycin selection to obtain stable integrants, individual clones were assessed for their response to Wnt-pathway activation. The clone with the highest dynamic range was used for the experiments depicted in Figures 2.3 and 2.4.

2.4.5 Cell culture and stimulation

HEK293T^{WOO} cells were cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12) supplemented with 10% FCS and 1% Penicillin/Streptomycin (GIBCO, Life Technologies) in 5% CO2 at 37 °C. These cells respond to activation of the Wnt/β-catenin signaling pathway by expressing firefly luciferase, since the firefly luciferase in the 7TFP construct is driven by the 7×Tcf promoter, which contains 7 repeats of the TCF/LEF transcription response element. Cells were plated the day prior to stimulation in a 96 well-plate at a density of 20.000 cells per well. Cells were stimulated with different concentrations (10-200 ng/ml) of purified Wnt3a protein (RnD) dissolved in 0.1% BSA in PBS, or with different concentrations (750 nM-6 μM) CHIR99021 (BioVision) dissolved in DMSO, for different amounts of time (1-24 hours). At the indicated time points following stimulation, cells were lysed in 20 μl of Passive Lysis Buffer (Promega) and cell lysate from the same experiment was used for both the luciferase assay (3 wells per condition) and Western blot analysis (the remainder of the 3 wells, pooled per lane).

2.4.6 Western blot analysis

Protein concentration was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Equal amounts of protein were run on an 8% SDS-PAGE gel. Proteins were transferred to a nitrocellulose membrane (Bio-Rad) and

blocked with TBS Odyssey Blocking Buffer (LI-COR Biosciences, diluted 1:1 in TBS prior to use). Primary antibodies directed against active β-catenin (Cat# 8814S, Cell Signaling, 1:1000), total β-catenin (Cat# 610153, BD Biosciences, 1:2000) and α-Tubulin (Cat# T9026, Sigma-Aldrich, 1:500) were diluted in blocking buffer supplemented with 0.1% Tween-20 (TBS-T). Staining was performed overnight at 4 °C. Membranes were washed in TBS-T followed by incubation with secondary antibodies (IRDye 680LT (Cat# 926-68021) or IRDye 800CW (Cat# 926-32212) (LI-COR), 1:20000 in TBS-T) for 2 hours. Membranes were washed in TBS-T and incubated in TBS prior to scanning at 700 nm and 800 nm using an Odyssey Fc (LI-COR Biosciences). Image Studio™ Lite 4.0 software (LI-COR Biosciences) was used to quantify relative protein levels. Background correction was performed according to the manufacturer's instructions (median of pixels, top/bottom border width of 3).

2.4.7 Luciferase assay

To measure the activity of firefly luciferase (and hence Wnt-pathway activity), 10 µl of cell lysate was transferred to a black 96-well Optiplate (Perkin Elmer). The SpectraMax L Microplate Luminometer was used to inject 50 µl Luciferase Assay Reagent II (Promega) per well followed by measurement of firefly luciferase activity.

References

Aberle H, Bauer A, Stappert J, Kispert A, Kemler R. β-catenin is a target for the ubiquitin–proteasome pathway. *EMBO J*. 1997, 16, 3797-804.

Albuquerque C, Bakker ER, van Veelen W, Smits R. Colorectal cancers choosing sides. *Biochim Biophys Acta*. 2011, 1816, 219-31.

Amit S, Hatzubai A, Birman Y, Andersen JS, Ben-Shushan E, Mann M, et al. Axin-mediated CKI phosphorylation of β-catenin at Ser 45: a molecular switch for the Wnt pathway. *Genes Dev*. 2002, 16, 1066-76.

Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer*. 2013, 13, 11-26.

Aulehla A, Herrmann BG. Segmentation in vertebrates: clock and gradient finally joined. *Genes Dev*. 2004, 18, 2060-7.

Behrens J, Jerchow B-A, Würtele M, Grimm J, Asbrand C, Wirtz R, et al. Functional Interaction of an Axin Homolog, Conductin, with β-Catenin, APC, and GSK3β. *Science*. 1998, 280, 596-99.

Behrens J, von Kries JP, Kühl M, Bruhn L, Wedlich D, Grosschedl R, et al. Functional interaction of β-catenin with the transcription factor LEF-1. *Nature*. 1996, 382, 638-42.

Benary U, Kofahl B, Hecht A, Wolf J. Modeling Wnt/beta-Catenin Target Gene Expression in APC and Wnt Gradients Under Wild Type and Mutant Conditions. *Front Physiol*. 2013, 4, 21.

Bhanot P, Brink M, Samos CH, Hsieh J-C, Wang Y, Macke JM, et al. A new member of the frizzled family from Drosophila functions as a Wingless receptor. *Nature*. 1996, 382, 225-30.

Bilić J, Huang Y-L, Davidson G, Zimmermann T, Cruciat C-M, Bienz M, et al. Wnt Induces LRP6 Signalosomes and Promotes Dishevelled-Dependent LRP6 Phosphorylation. *Science*. 2007, 316, 1619-22.

Bonzanni N, Garg A, Feenstra KA, Schutte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013, 29, i80-8.

Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. *Bioinformatics*. 2009, 25, 2049-56.

Buchert M, Athineos D, Abud HE, Burke ZD, Faux MC, Samuel MS, et al. Genetic Dissection of Differential Signaling Threshold Requirements for the Wnt/β-catenin Pathway In Vivo. *PLoS Genet*. 2010, 6, e1000816.

Cadigan KM, Peifer M. Wnt Signaling from Development to Disease: Insights from Model Systems. *Cold Spring Harb Perspect Biol*. 2009, 1, a002881.

Chia IV, Costantini F. Mouse Axin and Axin2/Conductin Proteins Are Functionally Equivalent In Vivo. *Mol Cell Biol*. 2005, 25, 4371-6.

Cho K-H, Baek S, Sung M-H. Wnt pathway mutations selected by optimal β-catenin signaling for tumorigenesis. *FEBS Letters*. 2006, 580, 3665-70.

Clevers H, Nusse R. Wnt/β-catenin Signaling and Disease. *Cell*. 2012, 149, 1192-205.

Davidson G, Wu W, Shen J, Bilic J, Fenger U, Stannek P, et al. Casein kinase 1γ couples Wnt receptor activation to cytoplasmic signal transduction. *Nature*. 2005, 438, 867-72.

de la Roche M, Worm J, Bienz M. The function of BCL9 in Wnt/beta-catenin signaling and colorectal cancer cells. *BMC Cancer*. 2008, 8, 199.

Fearon ER. Molecular Genetics of Colorectal Cancer. *Annu Rev Pathol*. 2011, 6, 479-507.

Fuerer C, R. N. Lentiviral Vectors to Probe and Manipulate the Wnt Signaling Pathway. *PLoS One*. 2010, 5, e9370.

Gonzalez A, Manosalva I, Liu T, Kageyama R. Control of Hes7 expression by Tbx6, the Wnt pathway and the chemical Gsk3 inhibitor LiCl in the mouse segmentation clock. *PLoS One*. 2013, 8, e53323.

Harris TJC, Peifer M. Decisions, decisions: β-catenin chooses between adhesion and transcription. *Trends Cell Biol*. 2005, 15, 234-7.

Henderson BR, Fagotto F. The ins and outs of APC and β-catenin nuclear transport. *EMBO Rep*. 2002, 3, 834-9.

Ikeda S, Kishida S, Yamamoto H, Murai H, Koyama S, Kikuchi A. Axin, a negative regulator of the Wnt signaling pathway, forms a complex with GSK-3β and β-catenin and promotes GSK-3β-dependent phosphorylation of β-catenin. *EMBO J*. 1998, 17, 1371-84.

Jamieson C, Sharma M, Henderson BR. Targeting the β-catenin nuclear transport pathway in cancer. *Seminars in Cancer Biology*. 2014, 27, 20-9.

Jho EH, Zhang T, Domon C, Joo CK, Freund JN, Costantini F. Wnt/β-Catenin/Tcf Signaling Induces the Transcription of Axin2, a Negative Regulator of the Signaling Pathway. *Mol Cell Biol*. 2002, 22, 1172-83.

Klaus A, Birchmeier W. Wnt signalling and its impact on development and cancer. *Nat Rev Cancer*. 2008, 8, 387-98.

Kofahl B, Wolf J. Mathematical modelling of Wnt/β-catenin signalling. *Biochem Soc Trans*. 2010, 38, 1281-5.

Kogan Y, Halevi-Tobias KE, Hochman G, Baczmanska AK, Leyns L, Agur Z. A new validated mathematical model of the Wnt signalling pathway predicts effective combinational therapy by sFRP and Dkk. *Biochemical Journal*. 2012, 444, 115-25.

Krepska E, Bonzanni N, Feenstra A, Fokkink W, Kielmann T, Bal H, et al. Design Issues for Qualitative Modelling of Biological Cells with Petri Nets. Proceedings of the Formal Methods in Systems Biology 2008, Vol 5054 of LNBI, Springer, Cambridge, UK. 2008, 48-62.

Lee E, Salic A, Krüger R, Heinrich R, Kirschner MW. The Roles of APC and Axin Derived from Experimental and Theoretical Analysis of the Wnt Pathway. *PLoS Biol*. 2003, 1, E10.

Leung JY, Kolligs FT, Wu R, Zhai Y, Kuick R, Hanash S, et al. Activation of AXIN2 expression by β-catenin-T cell factor. A feedback repressor pathway regulating Wnt signaling. *J Biol Chem*. 2002, 277, 21657-65.

Li VS, Ng SS, Boersema PJ, Low TY, Karthaus WR, Gerlach JP, et al. Wnt Signaling through Inhibition of β-catenin Degradation in an Intact Axin1 Complex. *Cell*. 2012, 149, 1245-56.

Li Y, Zhang Q, Yin X, Yang W, Du Y, Hou P, et al. Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Res*. 2011, 21, 196-204.

Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, Krishnadath KK, et al. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling. *Nat Genet*. 2000, 26, 146-7.

Lloyd-Lewis B, Fletcher AG, Dale TC, Byrne HM. Toward a quantitative understanding of the Wnt/β-catenin pathway through simulation and experiment. *Wiley Interdiscip Rev Syst Biol Med*. 2013, 5, 391-407.

Lustig B, Jerchow B, Sachs M, Weiler S, Pietsch T, Karsten U, et al. Negative Feedback Loop of Wnt Signaling through Upregulation of Conductin/Axin2 in Colorectal and Liver Tumors. *Mol Cell Biol*. 2002, 22, 1184-93.

MacDonald BT, Tamai K, He X. Wnt/β-catenin signaling: components, mechanisms, and diseases. *Dev Cell*. 2009, 17, 9-26.

Mazzoni SM, Fearon ER. AXIN1 and AXIN2 variants in gastrointestinal cancers. *Cancer Lett*. 2014, 355, 1-8.

Molenaar M, van de Wetering M, Oosterwegel M, Peterson-Maduro J, Godsave S, V. K, et al. XTcf-3 Transcription Factor Mediates β-Catenin-Induced Axis Formation in Xenopus Embryos. *Cell*. 1996, 86, 391-9.

Morin PJ, Sparks AB, Korinek V, Barker N, Clevers H, Vogelstein B, et al. Activation of β-Catenin-Tcf Signaling in Colon Cancer by Mutations in β-Catenin or APC. *Science*. 1997, 275, 1787-90.

Mosimann C, Hausmann G, Basler K. β-catenin hits chromatin: regulation of Wnt target gene activation. *Nat Rev Mol Cell Biol*. 2009, 10, 276-86.

Naik S, Piwnica-Worms D. Real-time imaging of β-catenin dynamics in cells and living mice. *Proc Natl Acad Sci U S A*. 2007, 104, 17465-70.

Nusse R. Howard Hughes Medical Institute and the Department of Developmental Biology at Stanford University, California, USA; 2016, Available: http://web.stanford.edu/group/nusselab/cgi-bin/wnt/.

Petri CA. Kommunikation mit Automaten. PhD Thesis, Technische Universität Darmstadt, Bonn, Germany. 1962.

Polakis P. Wnt Signaling in Cancer. *Cold Spring Harb Perspect Biol*. 2012, 4, a008052.

Reisig CA, Rozenberg G. Lectures on Petri Nets I: Basic Models, Vol. 1491 and 1492 of LNCS. Springer, Berlin, Germany. 1998.

Ring DB, Johnson KW, Henriksen EJ, Nuss JM, Goff D, Kinnick TR, et al. Selective Glycogen Synthase Kinase 3 Inhibitors Potentiate Insulin Activation of Glucose Transport and Utilization In Vitro and In Vivo. *Diabetes*. 2003, 52, 588-95.

Rubinfeld B, Robbins P, El-Gamil M, Albert I, Porfiri E, Polakis P. Stabilization of β-Catenin by Genetic Defects in Melanoma Cell Lines. *Science*. 1997, 275, 1790-2.

Ruths D, Muller M, Tseng J-T, Nakhleh L, Ram PT. The Signaling Petri Net-Based Simulator: A Non-Parametric Strategy for Characterizing the Dynamics of Cell-Specific Signaling Networks. *PloS CB*. 2008, 4, e1000005.

Satoh S, Daigo Y, Furukawa Y, Kato T, Miwa N, Nishiwaki T, et al. AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat Genet*. 2000, 24, 245-50.

Schwarz-Romond T, Metcalfe C, Bienz M. Dynamic recruitment of axin by Dishevelled protein assemblies. *J Cell Sci*. 2007, 120, 2402-12.

Stamos JL, Weis WI. The β-catenin Destruction Complex. *Cold Spring Harb Perspect Biol*. 2013, 5, a007898.

Suriben R, Fisher DA, Cheyette BN. Dact1 presomitic mesoderm expression oscillates in phase with Axin2 in the somitogenesis clock of mice. *Dev Dyn*. 2006, 235, 3177-83.

Tamai K, Semenov M, Kato Y, Spokony R, Liu C, Katsuyama Y, et al. LDL-receptor-related proteins in Wnt signal transduction. *Nature*. 2000, 407, 530-5.

Tan CW, Gardiner BS, Hirokawa Y, Layton MJ, Smith DW, Burgess AW. Wnt Signalling Pathway Parameters for Mammalian Cells. *PLoS One*. 2012, 7, e31882.

van de Wetering M, Cavallo R, Dooijes D, van Beest M, van Es J, Loureiro J, et al. Armadillo Coactivates Transcription Driven by Product of the Drosophila Segment Polarity Gene dTCF. *Cell*. 1997, 88, 789-99.

van Leeuwen IM, Byrne HM, Jensen OE, King JR. Elucidating the interactions between the adhesive and transcriptional functions of β-catenin in normal and cancerous cells. *J Theor Biol*. 2007, 247, 77-102.

Veeman MT, Slusarski DC, Kaykas A, Louie SH, Moon RT. Zebrafish Prickle, a Modulator of Noncanonical Wnt/Fz Signaling, Regulates Gastrulation Movements. *Current Biology*. 2003, 13, 680-5.

Verkaar F, Blankesteijn WM, Smits JF, Zaman GJ. β-Galactosidase enzyme fragment complementation for the measurement of Wnt/β-catenin signaling. *FASEB J*. 2010, 24,1205-17.

Verkaar F, Cadigan KM, van Amerongen R. Celebrating 30 Years of Wnt Signaling. *Sci Signal*. 2012, 5, mr2.

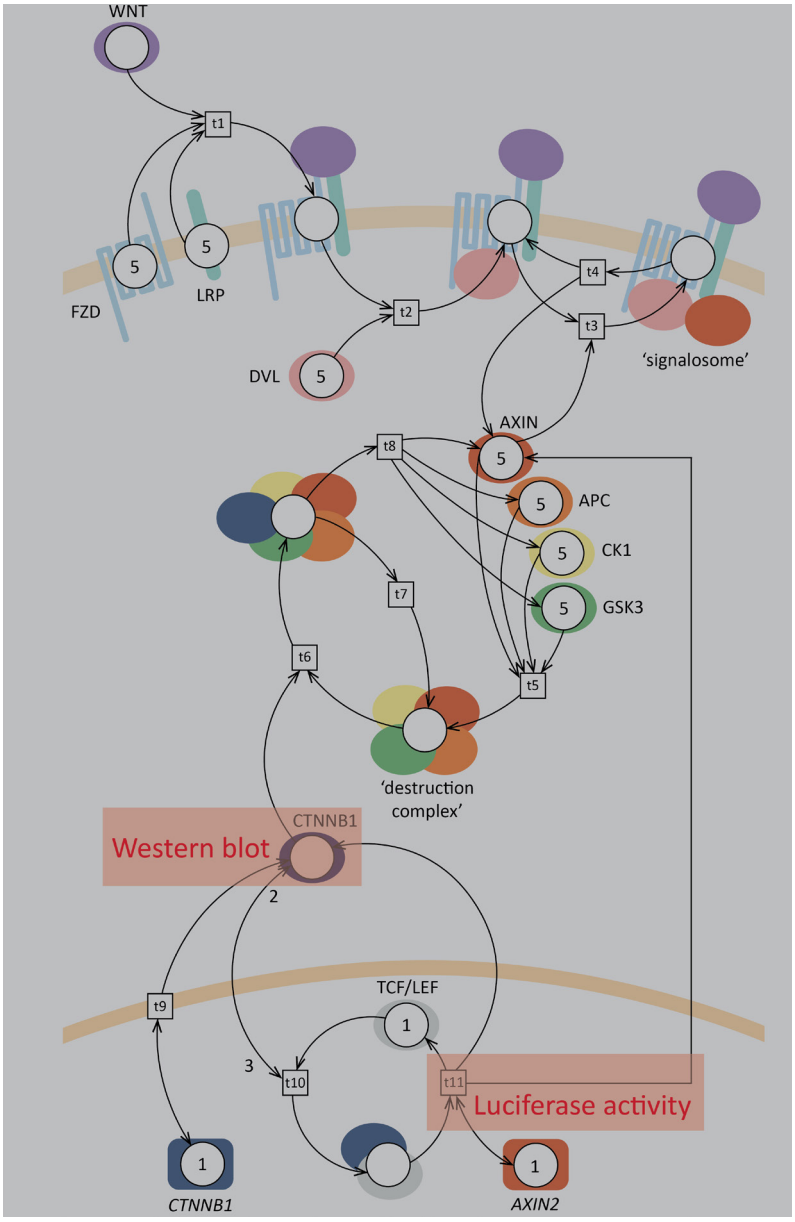
Wawra C, Kuhl M, Kestler HA. Extended analyses of the Wnt/β-catenin pathway: robustness and oscillatory behaviour. *FEBS Lett*. 2007, 581, 4043-8.

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, et al. The ground state of embryonic stem cell self-renewal. *Nature*. 2008, 453, 519-23.

Yokoyama N, Yin D, Malbon CC. Abundance, complexation, and trafficking of Wnt/β-catenin signaling elements in response to Wnt3a. *J Mol Signal*. 2007, 2, 11.

Zeng L, Fagotto F, Zhang T, Hsu W, Vasicek TJ, Perry WL, et al. The Mouse Fused Locus Encodes Axin, an Inhibitor of the Wnt Signaling Pathway That Regulates Embryonic Axis Formation. *Cell*. 1997, 90, 181-92.

Zeng X, Tamai K, Doble B, Li S, Huang H, Habas R, et al. A dual-kinase mechanism for Wnt co-receptor phosphorylation and activation. *Nature*. 2005, 438, 873-7.



S2.1 Figure. Illustration of the 'location' of the experimental readouts of Western blot analysis and luciferase reporter assay. The 'location' of the readouts of Western blot analysis and luciferase reporter assay compared to their equivalent implementation in our Petri net model. Western blot analysis measures the protein levels of β-catenin (i.e. measures the β-catenin accumulation), whereas the luciferase reporter assay measures the transcriptional activation (by β-catenin/TCF complexes) of the TOPFLASH reporter, which is followed by the production of luciferase protein. Since the luciferase reporter activation requires β-catenin to be present it takes longer for the TOPFLASH reporter to be activated than for the increase in β-catenin protein levels to occur i.e. the activation always occurs downstream of β-catenin accumulation.

CHAPTER 3

Aurora kinase A (AURKA) interaction with Wnt and Ras-MAPK signaling pathways in colorectal cancer

Sci. Rep. 2018, 8(7522). doi:10.1038/s41598-018-24982-z

Annika Jacobsen, Linda J.W. Bosch, Sanne R. Martens-de Kemp, Beatriz Carvalho, Anke H. Sillars-Hardebol, Richard J. Dobson, Emanuele de Rinaldis, Gerrit A. Meijer, Sanne Abeln, Jaap Heringa, Remond J.A. Fijneman, K. Anton Feenstra

Abstract

Hyperactivation of Wnt and Ras-MAPK signaling are common events in development of colorectal adenomas. Further progression from adenoma-to-carcinoma is frequently associated with 20q gain and overexpression of Aurora kinase A (AURKA). Interestingly, AURKA has been shown to further enhance Wnt and Ras-MAPK signaling. However, the molecular details of these interactions in driving colorectal carcinogenesis remain poorly understood.

Here we first performed differential expression analysis (DEA) of *AURKA* knockdown in two colorectal cancer (CRC) cell lines with 20q gain and AURKA overexpression. Next, using an exact algorithm, Heinz, we computed the largest connected protein-protein interaction (PPI) network module of significantly deregulated genes in the two CRC cell lines. The DEA and the Heinz analyses suggest 20 Wnt and Ras-MAPK signaling genes being deregulated by AURKA, whereof *β-catenin* and *KRAS* occurred in both cell lines. Finally, shortest path analysis over the PPI network revealed eight 'connecting genes' between AURKA and these Wnt and Ras-MAPK signaling genes, of which *UBE2D1*, *DICER1*, *CDK6* and *RACGAP1* occurred in both cell lines.

This study, first, confirms that AURKA influences deregulation of Wnt and Ras-MAPK signaling genes, and second, suggests mechanisms in CRC cell lines describing these interactions.

3.1 Introduction

Colorectal cancer (CRC) is the third most common cancer in men and the second most common cancer in women worldwide (Ferlay et al., 2015). In the early stages of CRC development, proliferative signaling is sustained by hyperactivation of the Wnt and Ras-MAPK signaling pathways due to mutations in key regulatory genes (Hanahan and Weinberg, 2011). Disruption of Wnt signaling, caused by mutations in the APC tumor suppressor gene or other genes such as *CTNNB1* (hereafter referred to as *β-catenin*), *AXIN1* or *AXIN2*, promotes the progression from normal colon epithelium to a benign precursor lesion, called adenoma (Fearon, 2011). Subsequently, adenoma-to-carcinoma progression is driven by further genetic and epigenetic alterations. For example, in addition to the Ras-MAPK pathway, activated by mutation in e.g. the *KRAS* gene, other pathways important in carcinoma development are the TGFβ pathway, disrupted by mutation in e.g. SMAD4, and the TP53 pathway, disrupted by mutations in the *TP53* gene (Haan et al., 2014; The Cancer Genome Atlas Network, 2012). Recently, it has been shown that adenoma organoids harboring all these mutations can induce invasive cancers in mice only when a background of chromosomal instability is present (Matano et al., 2015). This signifies the importance of chromosomal instability, which in fact occurs in ~85% of CRC (Lengauer et al., 1997), and is characterized by gross chromosomal aberrations.

Chromosomal arm 20q is frequently gained in CRC (Meijer et al., 1998; Nakao et al., 2004) and has a strong association with the progression of colorectal adenoma to carcinoma (Carvalho et al., 2009). Aurora kinase A (AURKA), a gene coding for a key cell cycle regulator, is located on 20q. There is a significant correlation between the 20q copy number and increased AURKA mRNA and protein expression (Sillars-Hardebol et al., 2012). Gain of 20q and/or AURKA overexpression is associated with a poor prognosis in many cancer types including CRC (Aust et al., 2004; Belt et al., 2012; Goos et al., 2013; Lassus et al., 2011; Tanner et al., 1995; Zeng et al., 2014).

AURKA overexpression has been shown to stabilize β-catenin levels and thereby activating Wnt signaling in gastric cancer cells by phosphorylating the negative regulator of β-catenin, GSK3B (Dar et al., 2009; Liu et al., 2016). Also in glioma-initiating cells (distinguished by their capacity of self-renewal) AURKA is a negative regulator of β-catenin, by binding to AXIN1 (Xia et al., 2013). Recently, it has been shown that AURKA upregulates Ras-MAPK signaling by interacting with the H-RAS/Raf-1 complex in kidney cells (Umstead et al., 2017). In addition, AURKA itself has been shown to be a target gene of both MAPK1/ERK2 signaling in pancreatic cancer cells (Furukawa et al., 2006) and Wnt/β-catenin signaling in multiple myeloma (Dutta-Simmons et al., 2009). These data suggest a positive feedback loop from hyperactive proliferative signaling to AURKA overexpression, further inducing proliferative signaling cells (Umstead et al., 2017).

All this implies that there is interplay between AURKA and the Wnt and Ras-MAPK signaling pathways and vice versa in different cancer settings. For Wnt and Ras-MAPK signaling, much of the mechanisms have been elucidated, also in relation with CRC (Jacobsen et al., 2016), but such detail is not available for the interplay with AURKA. Although different molecular mechanisms are observed in the different settings this argues that the regulation itself is important. In this study, we used two distinct cell lines, SW480 and Caco2, both derived from colon carcinomas with 20q copy number gain and mutated TP53. However, the genetics differ between these cell lines, they originate from different individuals and therefore have different germline variations, and they have progressed to carcinomas independently and therefore also differ in their somatic alterations, such as DNA mutations and DNA copy number alterations. The distinct DNA copy number profiles of the two cell lines are shown in S3.1 Figure. Further, a comparative study of colon cancer cell lines showed that out of five critical cancer genes (KRAS, BRAF, PIK3CA, PTEN, and TP53) only the mutation status of KRAS differed between the two cell lines (mutated in SW480). Further, of the other four genes TP53 was mutated in both cell lines (Ahmed et al., 2013). Using SW480 and Caco2, we set out to investigate which key players and molecular interactions are involved in the interplay between AURKA and the Wnt and Ras-MAPK pathways that may drive progression of CRC.

3.2. Results

3.2.1 Differential mRNA expression analysis

Differential expression analysis (DEA) was performed upon *AURKA* siRNA-directed downmodulation in two CRC cell lines with 20q gain and AURKA overexpression, SW480 and Caco2. *AURKA* was the most differentially downregulated gene in both cell lines (p value of 1.69×10^{-6} in SW480 and 2.28×10^{-7} in Caco2), indicating that the siRNA experiment was successful. The number of significantly expressed genes in response to AURKA downmodulation at a q value less than 0.05 was 2,057 and 3,606 in SW480 and Caco2, respectively. 924 genes were significantly deregulated in both cell lines, whereof 50 were deregulated in opposite directions. In our analysis, however, we applied a more stringent threshold to determine significantly deregulated genes including both a q value less than 0.05 and a fold change greater than 1.5 or less than -1.5. This resulted in a 292 and 154 significantly deregulated genes in SW480 and Caco2, respectively. Of the 292 genes significantly deregulated in SW480, 139 genes were upregulated and 153 downregulated (Figure 3.1A and Additional File 1). Of the 154 genes significantly deregulated in Caco2, 73 genes were upregulated and 81 downregulated (Figure 3.1B and Additional File 2). Fifty-four genes were significantly deregulated in both cell lines (Figure 3.1C and Table 3.1). Fifty-three of these 54 were deregulated in the same direction in both cell lines: 28 up in both and 25 down in both. The only gene deregulated in

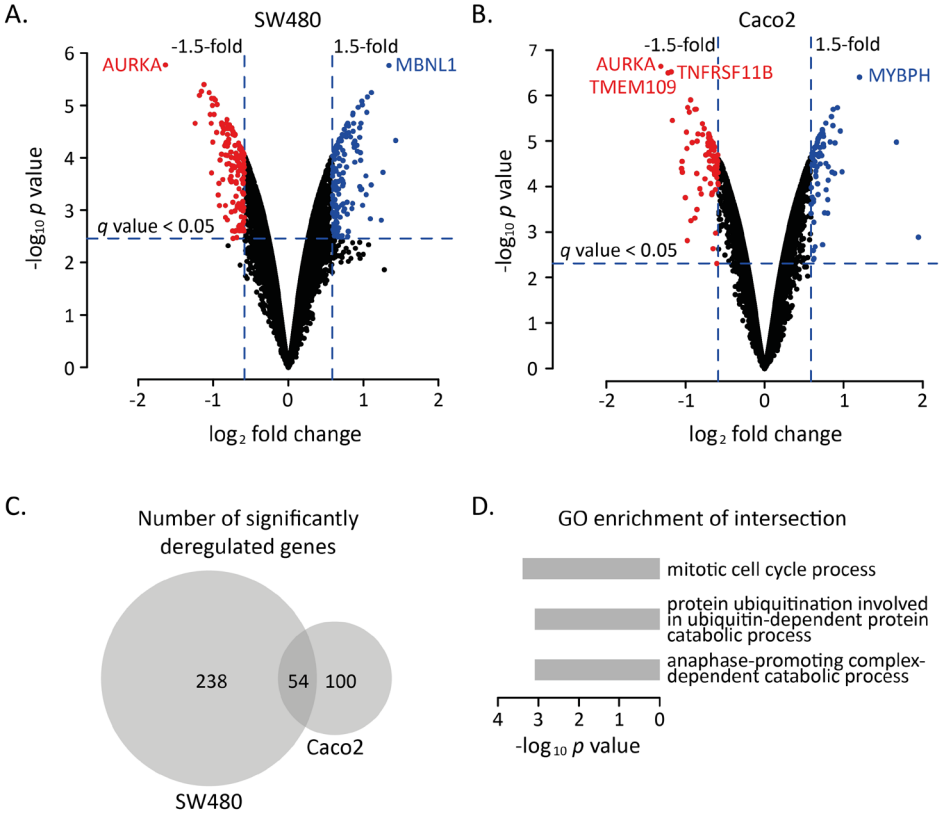


Figure 3.1. Differential mRNA expression analysis of *AURKA* knockdown in SW480 (A) and Caco2 (B) colorectal cancer cell lines. A) and B) Volcano plot of all unique genes from the differential mRNA expression analysis upon *AURKA* downmodulation for the SW480 and Caco2 cell lines, respectively. The vertical and horizontal lines illustrate the boundaries between significantly downregulated genes (red dots), significantly upregulated genes (blue dots) and non-significant differential genes (black dots). Horizontal lines represent the significance threshold of the *q* value < 0.05. Vertical lines represent the significance threshold of the fold change 1.5 (log_2 fold change 0.58) up or down. C) VENN diagram of the number of significantly deregulated genes in SW480 and Caco2 (see Table 1). D) GO enrichment of the genes (n=53) that are significantly deregulated in both cell lines in the same direction (the intersection in the Venn diagram in C minus one gene).

different directions was *SLC12A2*, which was up in Caco2 and down in SW480. These 53 genes are interesting candidates for the general influence of AURKA. GO enrichment analysis of these genes revealed that 'mitotic cell cycle' was the most significantly enriched process (Figure 31D).

3.2.2 Identification of the most significantly deregulated gene modules

In order to gain understanding from the sets of genes differentially expressed by *AURKA* knockdown, we applied the Heinz algorithm (Dittrich et al., 2008). Heinz integrates significant deregulation (the *p* values from the DEA described above) with

Table 3.1. Genes significantly deregulated in both the Caco2 and SW480 cell lines under AURKA knockdown (n=54). *p* value, *q* value and log_2 fold change (FC) from the two differential mRNA expression analyses are shown for each gene ID. The Gene names are retrieved from Ensembl.

Gene ID	Gene Name	SW480			Caco2		
		p value	*q* value	log_2FC	*p* value	*q* value	log_2FC
AURKA	Aurora kinase A	1.69×10^{-6}	1.86×10^{-2}	-1.64	2.28×10^{-7}	1.96×10^{-3}	-1.31
REEP5	Receptor accessory protein 5	2.20×10^{-5}	1.86×10^{-2}	-1.24	1.83×10^{-6}	5.18×10^{-3}	-0.97
SLC12A2	Solute carrier family 12 member 2	5.70×10^{-6}	1.86×10^{-2}	-1.05	1.39×10^{-5}	6.82×10^{-3}	0.70
DICER1	Dicer 1, ribonuclease III	1.94×10^{-4}	2.35×10^{-2}	-1.02	6.79×10^{-5}	1.00×10^{-2}	-0.75
FEZ2	Fasciculation and elongation protein zeta 2	7.41×10^{-6}	1.86×10^{-2}	-0.99	7.61×10^{-6}	6.82×10^{-3}	-0.70
KRAS	KRAS proto-oncogene, GTPase	9.58×10^{-6}	1.86×10^{-2}	-0.95	1.10×10^{-5}	6.82×10^{-3}	-0.65
UBE2D1	Ubiquitin conjugating enzyme E2 D1	3.08×10^{-4}	2.55×10^{-2}	-0.92	1.08×10^{-5}	6.82×10^{-3}	-0.91
TM4SF1	Transmembrane 4 L six family member 1	1.10×10^{-4}	2.28×10^{-2}	-0.91	2.03×10^{-5}	7.98×10^{-3}	-0.74
ACVR1	Activin A receptor type 1	2.60×10^{-4}	2.46×10^{-2}	-0.87	6.36×10^{-6}	6.82×10^{-3}	-0.99
MARVELD2	MARVEL domain containing 2	2.61×10^{-5}	1.86×10^{-2}	-0.87	2.13×10^{-5}	7.98×10^{-3}	-0.66
BORCS8	BLOC-1-related complex subunit 8	2.91×10^{-5}	1.86×10^{-2}	-0.86	1.61×10^{-5}	7.31×10^{-3}	-0.64
SLC25A24	Solute carrier family 25 member 24	5.87×10^{-5}	2.10×10^{-2}	-0.85	1.10×10^{-5}	6.82×10^{-3}	-0.71
MBOAT2	Membrane bound O-acyl-transferase domain containing 2	1.73×10^{-4}	2.35×10^{-2}	-0.76	7.00×10^{-6}	6.82×10^{-3}	-0.84
ACOX2	Acyl-CoA oxidase 2	3.59×10^{-5}	1.86×10^{-2}	-0.74	2.78×10^{-5}	8.77×10^{-3}	-0.61
TNFRSF11B	TNF receptor superfamily member 11b	3.13×10^{-5}	1.86×10^{-2}	-0.74	3.02×10^{-7}	1.96×10^{-3}	-1.18
NA*	ENSG00000260912*	3.47×10^{-3}	4.91×10^{-2}	-0.74	4.04×10^{-5}	9.30×10^{-3}	-1.05
NA*	ENSG00000234119*	3.74×10^{-5}	1.91×10^{-2}	-0.71	8.20×10^{-6}	6.82×10^{-3}	-0.69
SNX24	Sorting nexin 24	2.46×10^{-3}	4.32×10^{-2}	-0.70	5.16×10^{-5}	9.83×10^{-3}	-0.86
ANP32A	Acidic nuclear phosphoprotein 32 family member A	8.22×10^{-5}	2.26×10^{-2}	-0.70	2.15×10^{-5}	7.98×10^{-3}	-0.62
CANX	Calnexin	5.44×10^{-4}	2.89×10^{-2}	-0.70	2.32×10^{-6}	5.52×10^{-3}	-0.95
TMX2P1	Thioredoxin related transmembrane protein 2 pseudogene 1	2.87×10^{-4}	2.49×10^{-2}	-0.70	5.42×10^{-5}	9.96×10^{-3}	-0.63
B4GALT4	Beta-1,4-galactosyltransferase 4	8.17×10^{-4}	3.19×10^{-2}	-0.68	3.53×10^{-6}	6.82×10^{-3}	-1.17
SNX4	Sorting Nexin 4	1.09×10^{-4}	2.28×10^{-2}	-0.67	4.74×10^{-5}	9.65×10^{-3}	-0.61
TSPYL4	TSPY like 4	1.41×10^{-4}	2.35×10^{-2}	-0.66	8.55×10^{-6}	6.82×10^{-3}	-0.71
PHLDA2	Pleckstrin homology like domain family A member 2	1.67×10^{-3}	3.74×10^{-2}	-0.60	4.19×10^{-5}	9.37×10^{-3}	-0.61

(Table 3.1.continues on next page).

TP53I3	Tumor protein p53 inducible protein 3	1.31×10^{-4}	2.35×10^{-2}	-0.59	2.61×10^{-5}	8.77×10^{-3}	-0.60
ARHGAP19	Rho GTPase activating protein 19	2.25×10^{-4}	2.37×10^{-2}	0.59	7.93×10^{-6}	6.82×10^{-3}	0.74
LAMC1	Laminin subunit gamma 1	6.03×10^{-4}	2.95×10^{-2}	0.60	6.01×10^{-6}	6.82×10^{-3}	0.96
STK17B	Serine/threonine kinase 17b	1.32×10^{-3}	3.54×10^{-2}	0.61	5.71×10^{-5}	1.00×10^{-2}	0.65
NRP2	Neuropilin 2	2.83×10^{-3}	4.55×10^{-2}	0.65	4.83×10^{-5}	9.65×10^{-3}	0.85
IDNK	IDNK, gluconokinase	2.18×10^{-3}	4.17×10^{-2}	0.67	2.16×10^{-4}	1.34×10^{-2}	0.62
KTN1	Kinetin 1	4.50×10^{-5}	2.05×10^{-2}	0.68	1.88×10^{-5}	7.93×10^{-3}	0.60
RACGAP1	Rac GTPase activating protein 1	2.77×10^{-4}	2.49×10^{-2}	0.72	6.11×10^{-5}	1.00×10^{-2}	0.65
CDK6	Cyclin dependent kinase 6	8.63×10^{-5}	2.26×10^{-2}	0.75	4.60×10^{-6}	6.82×10^{-3}	0.86
SPA17	Sperm autoantigenic protein 17	1.51×10^{-4}	2.35×10^{-2}	0.77	2.00×10^{-5}	7.98×10^{-3}	0.66
LINC00467	Long intergenic non-protein coding RNA 467	3.30×10^{-5}	1.86×10^{-2}	0.80	8.62×10^{-6}	6.82×10^{-3}	0.68
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	3.99×10^{-5}	1.94×10^{-2}	0.83	1.84×10^{-5}	7.89×10^{-3}	0.70
BIK	BCL2 interacting killer	3.11×10^{-5}	1.86×10^{-2}	0.86	2.01×10^{-6}	5.18×10^{-3}	0.87
SH3D19	SH3 domain containing 19	1.22×10^{-4}	2.35×10^{-2}	0.89	1.05×10^{-5}	6.82×10^{-3}	0.83
EIF4EBP2	Eukaryotic translation initiation factor 4E protein binding 2	7.50×10^{-5}	2.26×10^{-2}	0.92	2.09×10^{-5}	7.98×10^{-3}	0.72
HMMR	Hyaluronan mediated motility receptor	4.07×10^{-5}	1.95×10^{-2}	0.95	1.17×10^{-5}	6.82×10^{-3}	0.74
KLHL15	Kelch like family member 15	3.55×10^{-4}	2.64×10^{-2}	0.95	2.46×10^{-4}	1.42×10^{-2}	0.59
DYRK2	Dual specificity tyrosine phosphorylation regulated kinase 2	4.09×10^{-4}	2.69×10^{-2}	0.97	3.68×10^{-5}	9.30×10^{-3}	0.80
ZNF268	Zink finger protein 268	3.95×10^{-4}	2.66×10^{-2}	0.98	8.72×10^{-5}	1.07×10^{-2}	0.67
NUP98	Nucleoporin 98	9.71×10^{-6}	1.86×10^{-2}	0.98	6.76×10^{-6}	6.82×10^{-3}	0.73
MTMR6	Myotubularin related protein 6	8.09×10^{-4}	3.19×10^{-2}	0.99	1.18×10^{-4}	1.20×10^{-2}	0.78
PTP4A1	Protein tyrosine phosphatase type IVA, member 1	5.08×10^{-4}	2.85×10^{-2}	1.04	1.94×10^{-4}	1.31×10^{-2}	0.80
GOLT1B	Golgi transport 1B	1.36×10^{-5}	1.86×10^{-2}	1.04	1.95×10^{-5}	7.98×10^{-3}	0.63
RPS27L	Ribosomal protein S27 like	6.91×10^{-6}	1.86×10^{-2}	1.05	1.06×10^{-5}	6.82×10^{-3}	0.67
MALAT1	Metastasis associated lung adenocarcinoma transcript 1	1.46×10^{-3}	3.66×10^{-2}	1.09	1.06×10^{-5}	6.82×10^{-3}	1.67
LAMP2	Lysosomal associated membrane protein 2	3.30×10^{-4}	2.62×10^{-2}	1.10	4.75×10^{-5}	9.65×10^{-3}	0.98
SYPL1	Synaptophysin like 1	5.69×10^{-6}	1.86×10^{-2}	1.11	5.63×10^{-6}	6.82×10^{-3}	0.73
ARRDC4	Arrestin domain containing 4	1.55×10^{-3}	3.70×10^{-2}	1.24	2.11×10^{-3}	3.15×10^{-2}	0.64
MBNL1	Muscleblind like splicing regulator 1	1.73×10^{-6}	1.86×10^{-2}	1.34	5.91×10^{-6}	6.82×10^{-3}	0.73

***Gene id and gene name are not available. The Ensembl id is provided as identifier.**

molecular protein interaction data (PPI network data from STRING) (Szklarczyk et al., 2015). The algorithm looks for significantly deregulated gene modules (sets of genes connected in the PPI network), which might include genes that show no expression changes, based on the intuition that some regulatory effects may involve changes that are not visible at the expression level. The genes in these modules can therefore be considered 'functionally' deregulated. First, weights are assigned to each node reflecting its *p* value from the DEA. The chosen FDR cut-off determines which nodes are assigned positive weights (*p* value below threshold) or negative (above threshold). The Heinz algorithm then computes the most significantly deregulated gene module, which is based both on the connectivity of the PPI network and the assigned weights. By thus combining the differential expression and protein interaction data, we can identify the most significantly deregulated gene modules (or functionally deregulated genes) upon *AURKA* downmodulation.

The FDR threshold was set to 6.21×10^{-4} for SW480 and 2.40×10^{-4} for Caco2. These thresholds were selected so that only the 50 most significantly deregulated genes were assigned positive weights. This selection allowed us to focus this analysis on gene modules of interpretable sizes (see Materials and Methods for further details). Figure 3.2 shows the most significantly deregulated gene modules for SW480 and Caco2. For SW480 the most significantly deregulated gene module consisted of 30 differentially expressed nodes (15 up and 15 down; see Figure 3.2A and S3.1 Table). Twenty-three were significantly deregulated (positive weight, circles in Figure 3.2A) and seven non-significant (negative weight, squares in Figure 3.2A). For Caco2 the most significantly deregulated gene module consisted of 31 differentially expressed nodes (13 up and 18 down; see Figure 3.2B and S3.2 Table). Twenty-nine were significantly deregulated (positive weight, circles in Figure 3.2B) and two non-significant (negative weight, square in Figure 3.2B. Note that remaining genes out of the 50 genes with positive weights do not occur in the largest connected module, because they are not connected in the PPI network. The intersection between the most significantly deregulated gene modules of each of the two cell lines contains nine genes: *AURKA*, *β-catenin*, *HMMR*, *KRAS*, *MBNL1*, *NUP98*, *REEP5*, *TNFRSF11B* and *UBE2D1* (black outline in nodes in Figure 3.2A-B). These are all directly connected in the modules (bold edges in Figure 3.2A-B), except for *MBNL1*, which is connected via one intermediate node: in SW480 via *JUN* to *KRAS*, *β-catenin* and *TNFRSF11B*, and in Caco2 via *ALB* to *KRAS*, *AURKA*, *β-catenin* and *TNFRSF11B*, and via *FOS* to *KRAS*, *UBE2D1*, *β-catenin* and *TNFRSF11B*.

This integrative approach, using the *p* values from the DEA and the PPI network data from STRING determined 30 and 31 genes in the most significantly deregulated gene modules for SW480 and Caco2, respectively. This analysis identified three genes for SW480 (*JUN*, *RPS3A* and *LCK*) and two genes for Caco2 (*β-catenin* and *FOS*), which were not significantly deregulated in the DEA.

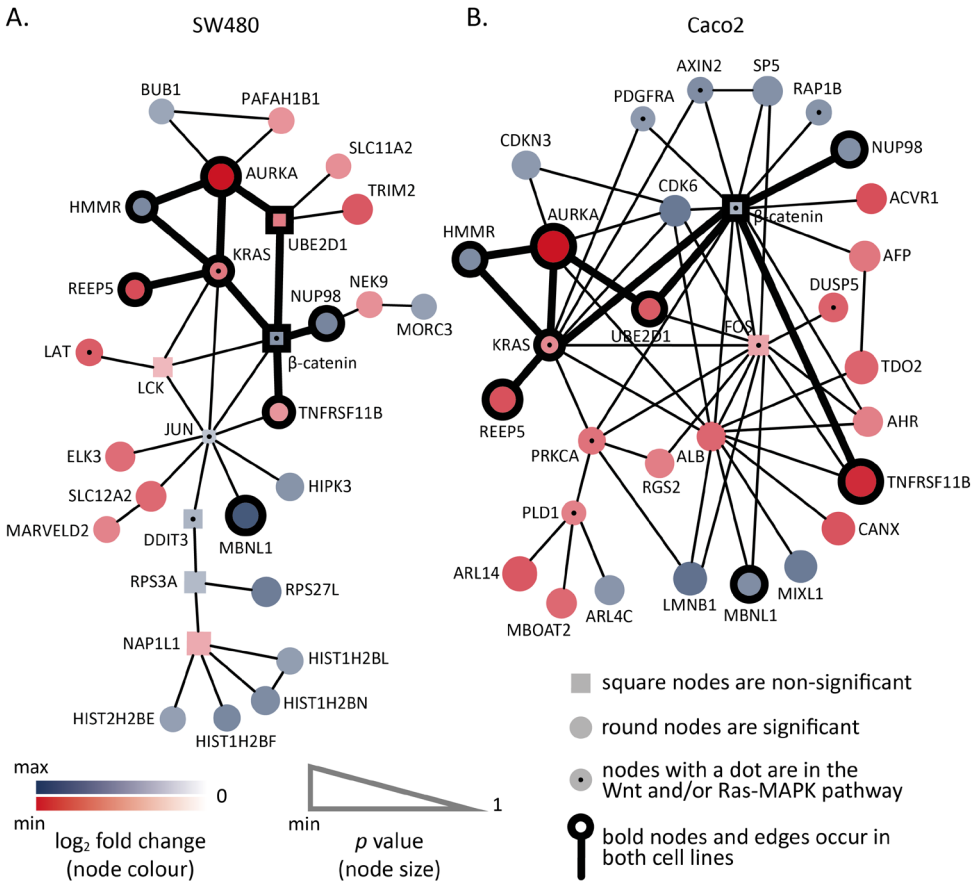


Figure 3.2. Significantly deregulated gene modules in SW480 (A) and Caco2 (B) derived from integrating *p* values from differential mRNA expression analysis (DEA) upon downmodulation of *AURKA* and STRING protein-protein interaction network data using the Heinz algorithm. A) and B) Based on the chosen FDR threshold of 6.21×10^{-4} for SW480 and 2.40×10^{-4} for Caco2 in the Heinz analysis, round nodes are significant (*p* value lower than the FDR threshold), and squared nodes are non-significant (*p* value above the FDR threshold). The size of the nodes is based on the *p* values from the DEA: a lower *p* value corresponds to a bigger node. The color of the nodes indicates the direction of the fold change with respect to *AURKA* downmodulation in the DEA. Red nodes are downregulated and blue nodes upregulated. The color intensity of the nodes is correlated with the magnitude of the fold change, where the greatest intensities represent the highest absolute fold change. The nodes of the nine genes occurring in both modules are outlined in black and the direct edges between these are in bold. The nodes representing proteins in the Wnt or Ras-MAPK pathways have a black dot.

3.2.3 Proteins and interactions involved in enhanced Wnt and Ras-MAPK signaling by AURKA

We then set out to annotate Wnt and Ras-MAPK signaling genes that were either significantly deregulated in the gene-centric DEA or found in the most significantly deregulated gene modules in the integrative network-aware Heinz analysis. In total, 20 genes in either of these two signaling pathways were identified: four genes in the Wnt pathway (*AXIN2*, *β-catenin*, *CTBP1* and *WNT5A*), eleven in the Ras-MAPK pathway (*DDIT3*, *PDGFRA*, *RAP1B*, *KITLG*, *FOS*, *DUSP5*, *PLD1*, *KRAS*, *LAT*, *MAP3K6*, RGL1), and five in both pathways (*PRKCA*, *PPP3CA*, *RAC2*, *PPP3R1* and *JUN*) (see Figure 3.3A). Two of these 20 deregulated genes, *KRAS* and *β-catenin*, appeared in both cell lines. The other 18 genes behaved cell line specific in these analyses: ten genes were found in SW480 only and eight genes were found only in Caco2 (Figure 3.3A). When looking at Wnt and Ras-MAPK genes compared to all selected genes by the DEA (significantly deregulated) and the Heinz analysis (part of the most significantly deregulated gene module), respectively, we see different enrichments. For the DEA this was 11/292 (4%) and 8/154 (5%) for SW480 and Caco2, respectively (colored nodes in Figure 3.3A, and totals in Figure 3.1C). On the other hand, for the Heinz analysis this was 5/30 (17%) and 9/31 (29%) for SW480 and Caco2, respectively (black dots vs. all nodes in Figure 3.2A-B). Hence, Wnt and Ras-MAPK signaling genes were much more enriched in the deregulated modules selected by the network-aware Heinz analysis compared to the DEA, suggesting that their mutual interaction within the associated PPI network modules is important for CRC progression.

The most significantly deregulated gene modules also suggest interactions between AURKA and the Wnt and Ras-MAPK pathway (Figure 3.2). Only KRAS is directly connected to AURKA, whereas the other genes are indirectly connected to AURKA via one or more genes. In order to identify interactions between AURKA and the remaining genes of the 20 Wnt and Ras-MAPK genes, we applied a shortest path analysis through the protein interaction network. This analysis was repeated for each of the cell lines, SW480 and Caco2. We used a subset of the STRING PPI only consisting of the genes that were either significantly deregulated in the DEA or part of the most significantly deregulated module from the Heinz analysis. The shortest path between each of these genes to AURKA was determined for SW480 (Figure 3.3B) and Caco2 (Figure 3.3C). Some of these interactions are directly from AURKA to Wnt and Ras-MAPK genes, but most are indirectly via additional interactions outside Wnt or Ras-MAPK signaling. For both cell lines there were six such external 'connecting genes': *LCK* and *PPP1CB* in SW480, and *ALB* and *CASK* in Caco2, and four shared: *UBE2D1*, *DICER1*, *CDK6* and *RACGAP1*. We propose that these genes could be important to explain the observed correlation between AURKA modulation and Wnt and Ras-MAPK activity.

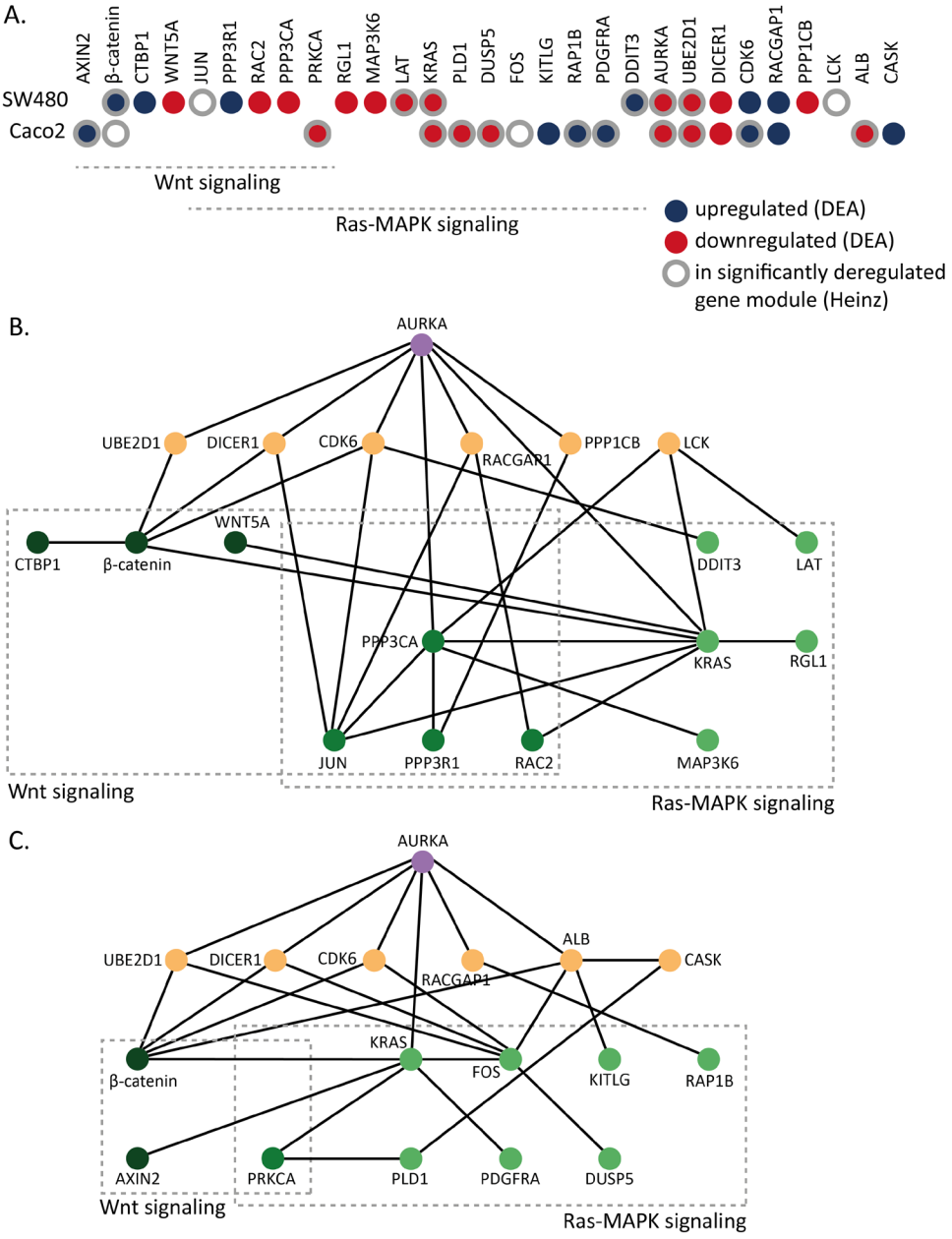


Figure 3.3. Wnt and Ras-MAPK pathway genes and interacting genes with AURKA, significantly deregulated (DEA) or in the most significantly deregulated gene module (Heinz) in the two cell lines, SW480 and Caco2. A) Overview of AURKA, the genes in the Wnt and Ras-MAPK pathways, and the genes connecting these, either significantly deregulated (DEA; *q* value < 0.05, red for down and blue for upregulation) or found in the most significantly deregulated gene module (Heinz, grey edges) in the two cell lines. B) and C) Interactions between AURKA (top-middle, purple) and the Wnt and Ras-MAPK signaling genes (green shades) deduced together with additional 'connecting genes' (orange) using a shortest path analysis in SW480 (B) and Caco2 (C).

3.2.4 TCGA CRC data analysis

The Wnt and Ras-MAPK pathway proteins and their molecular interactions with AURKA identified in this study are based on two CRC cell lines with 20q gain where *AURKA* is downmodulated. These results provide us with a better understanding of possible interconnections between AURKA and Wnt and Ras-MAPK signaling, however, to gain insight into the CRC progression it is necessary to study these genes from a cellular context that better represents the physiological environment. In CRC tissue samples where epithelial neoplastic cells interact with their tumor microenvironment the biology is more complex and heterogeneous than in isolated epithelial cancer cell lines.

We analyzed publicly available CRC samples from TCGA (RNA-seq and somatic copy number aberration data) to see if we could recover the significantly deregulated genes from the CRC cell line DEA. Out of 330 microsatellite stable tumor samples, 217 (65.76%) had *AURKA* copy number aberration, determined by a segment mean threshold of 0.4 in the somatic copy number aberration data (see Materials and Methods). DEA was performed between the *AURKA*-no-gain and *AURKA*-gain samples using the RNA-seq expression data. Note this atypical order of the sample groups (no-gain vs. gain instead of gain vs. no-gain) is to ensure proper comparison to the distinctive setup of the cell line experiments where *AURKA* is downmodulated (see Materials and Methods). *AURKA* was significantly deregulated (p value = 1.62×10^{-26}), demonstrating the correlation between increased *AURKA* count and copy number gain. In total, 1308 genes were significantly deregulated in the *AURKA*-no-gain samples with a significance threshold of the adjusted p value less than 10-5, 751 up and 557 down (Additional File 3). We then investigated the significance and directionality of the genes located on 20q. First, we retrieved the Ensembl ids for the genes located on 20q11-20q13.33 from BioMart. Second, we mapped these Ensembl ids to the Ensembl ids in the TCGA samples, which resulted in 316 genes. Out of these, 168 were significantly deregulated, 164 down and 4 up. Thus, 163 genes located on 20q11-20q13.33 are significantly deregulated in the same direction as AURKA (Additional File 4).

We compared the significantly deregulated genes in the TCGA *AURKA*-no-gain samples (n = 1308) to the sets of genes significantly deregulated in the SW480 (n = 292) and Caco2 (n = 154) DEA (described above). A minor overlap was found; only 24 genes were significantly deregulated in the TCGA and either of the cell lines (14 in SW480 and 13 in Caco2). Further, the majority of these genes were deregulated in opposite directions between the two different settings (see S3.2 Figure).

3.3. Discussion

In this study we identified 20 Wnt and Ras-MAPK signaling genes and eight additional connecting genes that suggest possible mechanisms of interaction between AURKA and Wnt and Ras-MAPK signaling in a CRC cell line setting. We performed mRNA expression analysis on AURKA knockdown in two CRC

cell lines, SW480 and Caco2, and applied three main computational approaches to interpret this expression data. First, we applied straightforward gene-centric DEA determining significantly deregulated genes. Second, we applied the Heinz algorithm, an integrative network-aware analysis, using the expression data and the STRING PPI network data to determine significantly deregulated gene modules. Third, we applied a shortest path analysis determining connections between AURKA and Wnt and Ras-MAPK signaling genes that were significantly deregulated in the first two analyses, DEA and Heinz. The DEA resulted in the largest set of genes, however it does not by itself reveal how their deregulation may be intertwined. The Heinz integrative analysis clearly revealed in an unbiased way the involvement of the Wnt and Ras-MAPK pathways, as well as indicated some plausible mechanisms of interaction between AURKA and both pathways. Finally, the shortest path analysis added 8 connecting genes that may be relevant to explain how *AURKA* modulation may affect the activity of these pathways, and thereby further induce proliferation in developing carcinoma cells.

In previous studies interactions between AURKA and Wnt signaling have been observed in gastric cancer cell lines (Dar et al., 2009; Liu et al., 2016) and glioma-initiating cells (Xia et al., 2013), and interactions between AURKA and Ras-MAPK signaling have been observed in kidney cells (Umstead et al., 2017), pancreatic cancer cells (Furukawa et al., 2006) and multiple myeloma (Dutta-Simmons et al., 2009). Thus the link between AURKA and Wnt and Ras-MAPK signaling is not new in general, but the specific proteins that are deregulated differ between the different cellular contexts. For instance, GSK3B and AXIN1 are two proteins deregulated by AURKA leading to stimulation of Wnt signaling in gastric cancer cell lines (Dar et al., 2009; Liu et al., 2016) and glioma-initiating cells (Xia et al., 2013), respectively. Activation of Ras-MAPK by AURKA has been attributed to interaction between AURKA and the H-RAS/Raf-1 complex in kidney cells (Umstead et al., 2017). In our analysis, in the two CRC cell lines, SW480 and Caco2, differential expression of *GSK3B*, *AXIN1* and *HRAS*, was not recovered, however, *AXIN2* and *KRAS*, homologous of *AXIN1* and *HRAS*, respectively, were recovered. Besides the differences in the cellular contexts, a plausible explanation for not recovering *GSK3B* and *AXIN1* in our analysis could be that GSK3B phosphorylation and AXIN1 binding to AURKA and β-catenin involve changes at the protein level, which would not be visible in our mRNA expression data. Another limitation of our experimental setup is that *AURKA* is the only gene that is downmodulated, whereas in an in vivo setting more genes on 20q are deregulated together with *AURKA*. One example is *TPX2*, which just as with *AURKA* is located on 20q, and implied to be involved with gain of 20q (Sillars-Hardebol et al., 2012), however, in our analysis TPX2 is not modulated.

Based on our analysis, the genes in the Wnt and Ras-MAPK pathways affected by *AURKA* downmodulation were mostly unique per cell line. In the Wnt pathway, only *β-catenin* was affected in both cell lines, and in the Ras-MAPK pathway, it was only *KRAS*. It should be noted, however, that these genes are at the core of

their respective pathways, which suggests that in both cellular contexts, perturbing these pathways may be very important. The molecular differences can be explained by the cause or the effect of the perturbations being dependent on the particular cellular context. On the other hand, the 'connecting genes' (connecting AURKA to the Wnt and Ras-MAPK signaling genes) were mostly similar between the cell lines (four out of six genes shared in both). While both SW480 and Caco2 are CRC cell lines with 20q gain, AURKA overexpression, and TP53 mutation, they also show differences. Firstly, the cell lines originate from different individuals, and therefore different germline variations. Secondly, the two cell lines have progressed to carcinomas independently. One specific difference is that *KRAS* is mutated in SW480 but not in Caco2 (Ahmed et al., 2013). However, their genetic differences are much more extensive than this (see DNA copy number profiles in S3.1 Figure). Consequently, the underlying molecular biology of SW480 and Caco2 differs and they are therefore expected to respond differently to a perturbation such as *AURKA* downmodulation when looking at the expression of individual genes, although the overall changes in pathways could still be the same. For instance, KRAS is one of the most connected genes in both cell lines in our two network aware analyses (Figure 3.2 and 3.3), while at the same time the differentially expressed networks upon AURKA knockdown show different genes connected to KRAS in both cell lines. This could potentially be due to the different mutation status of KRAS between the cell lines. For instance, in the shortest path analysis (Figure 3.3) KRAS is connected to AURKA and β-catenin in both cell lines. However, in SW480, where KRAS is mutated, KRAS in addition is connected to LCK, WNT5A, PPP3CA, JUN, and RAC2. Whereas, in Caco2, KRAS is in addition connected to FOS, AXIN2, and PRKCA.

The four consensus molecular subtypes (CMS) of colorectal cancer distinguish differences in signaling activities (Guinney et al., 2015). CMS2 and CMS4 are the most common chromosomal instable subtypes, both containing 20q gain, with CMS2 having a better prognosis than CMS4. CMS2 show increased proliferation, whereas CMS4 exhibit invasion. In a recent study (Ubink et al., 2017), they found that four genes were sufficient to predict CMS4: *PDGFRA*, *PDGFRB*, *PDGFC* and *KIT*. Interestingly, in the current study, PDGFRA (p value = 1.28×10^{-5}) and the *KIT* ligand, *KITLG* (p value = 3.19×10^{-5}), were significantly deregulated in Caco2, but not in SW480 (possibly explained by the different genetic backgrounds). Thus upon *AURKA* downmodulation Caco2 resembles the CMS4 subtype, and it can be implied that Caco2 also displays the CMS4 phenotype, invasion.

The cell line results presented in this paper suggest possible interconnections between AURKA and Wnt and Ras-MAPK signaling. Analysis of the CRC tissue samples from the TCGA data provided results from a more appropriate cellular context for studying CRC progression. There was an overlap of 24 genes when comparing the significantly deregulated genes from the TCGA analysis and the cell line analyses. However, most of these overlapping significantly deregulated genes were deregulated in opposite directions between the two different settings (S3.2 Figure).

We should here take into account the big differences between these two settings. First, in the tissue samples from TCGA *AURKA* is gained as a result of gain in 20q, on which many more genes than only *AURKA* are located and affected (Additional File 4). In the cell line experiments, however, only *AURKA* is downmodulated, whereas all the other genes on 20q have remained unchanged. Second, in a tumor, epithelial neoplastic cells interact with their tumor microenvironment. Thus, the tissue samples are much more complex than the cell lines, and most likely there would be feedback mechanisms active in this environment that do not operate in the isolated cell lines. Third, a tumor is very heterogeneous, whereas a cell line is much more homogeneous. Thus, the presence of multiple clones within a tumor might also mask the effect of *AURKA*-gain present in a minority of the clones in the tumor. Thus, we believe that because of these big differences between the cell line and tissue settings, a gene that is deregulated in both cases does not necessarily have to be deregulated in the same direction, but the overlap of deregulated genes, and their pathways, provides a good starting point for further investigation of the mechanisms between AURKA and Wnt and Ras-MAPK signaling. At the same time, this also underlines the importance of cellular context when studying regulation of cancer.

In summary, gene-centric and network-aware analysis of CRC cell lines with *AURKA* knockdown shows that 20 Wnt and Ras-MAPK signaling genes, occurring in connected PPI networks in two CRC cell lines, are significantly deregulated by AURKA. The responses in the Wnt and Ras-MAPK pathways were different in the two cell lines; only *KRAS* (Ras-MAPK) and *β-catenin* (Wnt) were deregulated in both. Further, in both cell lines four genes connected *AURKA* to the deregulated genes in the Wnt and Ras-MAPK pathways: *CDK6*, *UBE2D1*, *DICER1* and *RACGAP1*. These results suggest possible genes and mechanisms for the interplay between AURKA and Wnt and Ras-MAPK signaling that are at the same time generic and unique between different CRC settings. Further investigation of the importance of Wnt and Ras-MAPK enhanced signaling by AURKA in CRC and the role of these genes for these interactions, will lead to a better understanding of the molecular mechanisms underlying CRC progression.

3.4. Materials and methods

3.4.1 Cell culture and transfection with small interfering RNA (siRNA)

SW480 cells were grown in Dulbecco's modified Eagle's medium (DMEM; Lonza, Verviers, Belgium) containing 10% fetal bovine serum (FBS) (HyClone; Perbio Science, Etten-Leur, The Netherlands), and Caco2 cells were grown in RPMI 1640 (Lonza) containing 20% FBS. Both cell culture media were supplemented with 2 mM L-glutamine, 100 IU/ml sodium penicillin (Astellas Pharma B.V., Leiderdorp, the Netherlands), and 100 mg/ml streptomycin (Fisiopharma, Palomonta (SA), Italy). Transfection with siRNA pools (SMARTpools) from Dharmacon (Lafayette,

Colorado, USA) was performed 24h after seeding according to the manufacturer's recommendations. A final siRNA concentration of 30 nM was obtained using DharmaFECT3 reagent (1:1000 dilution) for both cell lines. A non-targeting control siRNA pool (Non-Targeting Pool 2; D-001206-14) was used as a negative control.

3.4.2 RNA isolation

Total RNA was isolated using TriZol reagent (Invitrogen, Breda, the Netherlands) and subjected to purification using RNeasy Kit (Qiagen, Hilden, Germany). RNA concentrations and purities were measured on a Nanodrop ND-1000 spectrophotometer (Isogen, IJsselstein, the Netherlands). RNA quality was evaluated by generating an electropherogram on the Agilent Bioanalyzer 2100 using a RNA 6000 Nano-LabChip (Agilent Technologies, Santa Clara, CA, the Netherland). RNA integrity numbers (RIN) of > 9.0 were considered as good quality RNA.

3.4.3 Microarray expression analyses

Microarray expression experiments were performed on 4×44K Agilent expression arrays (Agilent 60-mer SurePrint technologies), as described before (Buffart et al., 2009). Two biological replicates were performed for each siRNA experiment (both AURKA and non-targeting). Pre-processing and differential expression analysis (DEA) was done using the R-Bioconductor packages Limma (Ritchie et al., 2015) and SVA-combat (Leek and Storey, 2007). In short, conventional background correction was applied, followed by within-array normalization using Loess. Subsequently, batch-effect removal was applied for the replicate experiments, after which between-array normalization was done using the quantile method. The microarray data have been submitted to NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE108320. DEA between AURKA siRNA transfected cells and non-targeted siRNA control transfected cells was assessed by using linear modeling and empirical Bayes statistics. Finally, p values were adjusted for multiple testing using conventional Benjamini Hochberg FDR correction. The threshold for significantly expressed genes was set at a q value less than 0.05 and a fold change greater than 1.5 or less than -1.5. The significantly deregulated genes are available in Additional File 1 for SW480 and Additional File 2 for Caco2. GO enrichment analysis was performed using the 'Gene Ontology enRIchment anaLysis and visuaLizAtion tool' (http://cbl-gorilla.cs.technion.ac.il/).

3.4.4 Probe ID mapping

The Agilent probe ids from the microarray expression experiments were mapped to their respective Ensembl ids and HGNC ids (when available) using BioMart mappings ('Agilent WholeGenome 4x44k v1 probe') downloaded from www.ensembl.org/biomart/martview/ (January 2017). Additional mapping of HGNC ids to Ensembl ids was done using mappings from the Agilent website http://www.chem.agilent.com/cag/bsp/gene_lists.asp (Human Genome, Whole - Four-Plex, 44K) (January 2017). We manually checked and mapped the significantly expressed

genes for each cell line. Out of the 30,889 and 30,105 Agilent probe ids 8,270 and 8,149 could not be mapped to Ensembl ids for Caco2 and SW480, respectively. Where the same Ensembl id was mapped to multiple Agilent probe ids, the Agilent probe id with the lowest *p* value was chosen. This filtering step resulted in 24,589 and 24,088 probes whereof 16,319 and 15,939 has unique Ensembl ids for Caco2 and SW480, respectively.

3.4.5 Protein-protein interaction network

The *Homo sapiens* protein-protein interaction (PPI) network used in this analysis was retrieved from the STRING database (medium confidence: 0.400) (Szklarczyk et al., 2015) (downloaded: January 2017). The proteins were mapped to Ensembl ids using BioMart mappings (as described in the previous section). Self- and duplicate interactions were removed. The network consisted of 680,790 PPIs.

To describe the nodes (genes) and edges (interactions), one node-file and one edge-file were constructed for each cell line, SW480 and Caco2, based on the intersection of nodes (Ensembl ids) in the *AURKA* knockdown DEA data and the STRING PPI data. The edges-files consisted of two columns with Ensembl ids representing the two interacting proteins. The node-file consisted of three columns: nodes (Ensembl id), *p* values, and $\log_2$ fold change. The Caco2 edge-file consisted of 456,552 edges, and the SW480 edge-file consisted of 437,865 edges. The Caco2 node-file consisted of 12,249 nodes, while that for SW480 consisted of 12,082 nodes.

3.4.6 Identification of the most significantly deregulated gene modules

The Heinz algorithm (Dittrich et al., 2008) was applied to determine the most significantly deregulated gene module (sets of genes connected in the PPI network) for each cell line, SW480 and Caco2, using the constructed edge- and node-files. Each node in the network was assigned a weight reflecting its *p* value from the DEA. A beta-uniform mixture model was fitted to the *p* value distribution determining its parameters: the shape parameter, α, and the mixture parameter, λ (Beisser et al., 2010). The α parameter was 0.342 for SW480 and 0.329 for Caco2. Due to the high amount of low *p* values, the λ parameter was approximated 0.0 in both cell lines. Before calculating the weights of the nodes we set the λ parameters to 0.1, which still gave a very reasonable fit. The Heinz algorithm then assigned weights to the nodes using the λ and α parameters, the *p* values and a false-discovery rate (FDR) threshold. Positive weights were given to nodes with a *p* value lower than the FDR threshold, whereas negative weights were given to nodes with a *p* value greater than the FDR threshold. The FDR threshold for SW480 was 6.21×10^{-4} corresponding to a *p* value of 4.46×10^{-5}. The FDR threshold for Caco2 was 2.40×10^{-4} corresponding to a *p* value of 1.37×10^{-5}. The Heinz algorithm subsequently calculated the maximum-scoring subnetwork of the STRING PPI network that represents the most significantly deregulated gene module. The genes in these modules and their *p* values and $\log_2$ fold changes are available in S3.1 Table for SW480 and S3.2 Table for Caco2.

Since the FDR threshold has an influence on how many nodes are assigned positive weights, it also has an influence on the size of the significantly deregulated gene module. We selected the FDR thresholds so that 50 genes had a positive weight for each cell line. These thresholds are a conservative choice to generate focus on the part of the network that shows the strongest deregulation. This means that genes with p values just above the threshold may still be significantly deregulated. The resulting modules selected by Heinz are not very sensitive to the FDR setting, simply said the modules just incrementally grow or shrink, as the threshold is adapted.

3.4.7 Determining Wnt and Ras-MAPK signaling genes

To define which genes are in the Wnt and Ras-MAPK pathways we downloaded the gene lists of the 'Wnt signaling pathway', 'MAPK signaling pathway' and 'Ras signaling pathway' from the KEGG database (January 2017) (Kanehisa et al., 2017). These genes were used to annotate Wnt and Ras-MAPK signaling genes in the results.

3.4.8 Shortest path analysis

To find possible connections between AURKA and the signaling pathways of interest, we performed a shortest-path analysis. The shortest path is defined as the minimum number of edges required to travel from one node in the PPI graph to another. First, we reduced the STRING PPI network to a subnetwork only consisting of the significantly deregulated genes determined by the DEA and the genes located in the most significantly deregulated gene module in the Heinz analysis. We then applied the Python package NetworkX to compute all the shortest paths between AURKA and the Wnt and Ras-MAPK signaling genes, where the edges were unweighted.

3.4.9 TCGA CRC data analysis

To compare the effect of AURKA on Wnt and Ras-MAPK signaling in CRC tissue data to our CRC cell line results we analyzed the TCGA COADREAD (The Cancer Genome Atlas Network, 2012) RNA-seq and somatic copy number aberration data for 330 microsatellite stable tumor samples retrieved from Firehose (March 2017). First, the samples were labeled by *AURKA*-gain and *AURKA*-no-gain based on the copy number aberration data in the genome region 54,944,445-54,967,393 (genome build 19) encoding the *AURKA* gene. Samples with a segment mean greater than or equal to 0.4 were labeled *AURKA*-gain, whereas samples with a segment mean less than 0.4 were labeled *AURKA*-no-gain (Additional File 5). The segment mean is the average of the tumor versus normal intensity $\log_2$ ratio, thereby describing the copy number aberrations. We selected the threshold of 0.4 following the assumption that the gain of *AURKA* also increases its expression (S3.3 Figure). Second, DEA was performed between the *AURKA*-no-gain and *AURKA*-gain samples. The atypical choice of performing a no-gain vs. gain comparison (instead of gain vs. no-gain) is because of the distinctive setup of the cell line experiments

themselves, where the non-perturbed cell lines have *AURKA*-gain and the perturbed cell lines have an *AURKA* downmodulation (no-gain). Significantly expressed genes based on the DEA were determined at an adjusted p value less than 10^{-5} (S3.4 Figure). The significantly deregulated genes are available in Additional File 3. We then investigated the significance and directionality of the genes located on 20q. First, we retrieved the Ensembl ids for the genes located on 20q11-20q13.33 from BioMart. Second, we mapped these Ensembl ids to the Ensembl ids in the TCGA samples. The TCGA DEA result for these genes is available in Additional File 4. Finally, we determined the overlap of significantly deregulated genes between the TCGA DEA and the DEA of SW480 and Caco2 (S3.2 Figure).

References

Ahmed D, Eide PW, Eilertsen IA, Danielsen SA, Eknaes M, Hektoen M, et al. Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis*. 2013, 2, e71.

Aust DE, Muders M, Kohler A, Schmidt M, Diebold J, Müller C, et al. Prognostic relevance of 20q13 gains in sporadic colorectal cancers: a FISH analysis. *Scand J Gastroenterol*. 2004, 39, 766-72.

Beisser D, Klau GW, Dandekar T, Müller T, Dittrich MT. BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics*. 2010, 26, 1129-30.

Belt EJ, te Velde EA, Krijgsman O, Brosens RP, Tijssen M, van Essen HF, et al. High lymph node yield is related to microsatellite instability in colon cancer. *Ann Surg Oncol*. 2012, 19, 1222-30.

Buffart TE, Tijssen M, El-Bchiri J, Duval A, van de Wiel MA, Ylstra B, et al. NMD inhibition fails to identify tumour suppressor genes in microsatellite stable gastric cancer cell lines. *BMC Med Genomics*. 2009, 2, 39.

Carvalho B, Postma C, Mongera S, Hopmans E, Diskin S, van de Wiel MA, et al. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut*. 2009, 58, 79-89.

Dar AA, Belkhiri A, El-Rifai W. The aurora kinase A regulates GSK-3β in gastric cancer cells. *Oncogene*. 2009, 28, 866-75.

Dittrich MT, Klau GW, Rosenwald A, Dandekar T, Muller T. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics*. 2008, 24, i223-i31.

Dutta-Simmons J, Zhang Y, Gorgun G, Gatt M, Mani M, Hideshima T, et al. Aurora kinase A is a target of Wnt/β-catenin involved in multiple myeloma disease progression. *Blood*. 2009, 114, 2699-708.

Fearon ER. Molecular genetics of colorectal cancer. *Annu Rev Pathol*. 2011, 6, 479-507.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015, 136, E359-E86.

Furukawa T, Kanai N, Shiwaku HO, Soga N, Uehara A, Horii A. AURKA is one of the downstream targets of MAPK1/ERK2 in pancreatic cancer. *Oncogene*. 2006, 25, 4831-9.

Goos JACM, Coupe VMH, Diosdado B, Delis-Van Diemen PM, Karga C, Belien JAM, et al. Aurora kinase A (AURKA) expression in colorectal cancer liver metastasis is associated with poor prognosis. *Br J Cancer*. 2013, 109, 2445-52.

Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015, 21, 1350-6.

Haan JC, Labots M, Rausch C, Koopman M, Tol J, Mekenkamp LJ, et al. Genomic landscape of metastatic colorectal cancer. *Nat Commun*. 2014, 5, 5457.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011, 144, 646-74.

Jacobsen A, Heijmans N, Verkaar F, Smit MJ, Heringa J, van Amerongen R, et al. Construction and Experimental Validation of a Petri Net Model of Wnt/β-Catenin Signaling. *PLoS One*. 2016, 11, e0155743. **Chapter 2 of this thesis.**

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017, 45, D353-D61.

Lassus H, Staff S, Leminen A, Isola J, Butzow R. Aurora-A overexpression and aneuploidy predict poor outcome in serous ovarian carcinoma. *Gynecol Oncol*. 2011, 120, 11-7.

Leek JT, Storey JD. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genetics*. 2007, 3, e161.

Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. *Nature*. 1997, 386, 623-7.

Liu X, Li Z, Song Y, Wang R, Han L, Wang Q, et al. AURKA induces EMT by regulating histone modification through Wnt/β-catenin and PI3K/Akt signaling pathway in gastric cancer. *Oncotarget*. 2016, 7, 33152-64.

Matano M, Date S, Shimokawa M, Takano A, Fujii M, Ohta Y, et al. Modeling colorectal cancer using CRISPR-Cas9–mediated engineering of human intestinal organoids. *Nat Med*. 2015, 21, 256-62.

Meijer GA, Hermsen MAJA, Baak JPA, van Diest PJ, Meuwissen SGM, Beliën JAM, et al. Progression from colorectal adenoma to carcinoma is associated with non-random chromosomal gains as detected by comparative genomic hybridisation. *J Clin Pathol*. 1998, 51, 901-9.

Nakao K, Mehta KR, Fridlyand J, Moore DH, Jain AN, Lafuente A, et al. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis*. 2004, 25, 1345-57.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015, 43, e47.

Sillars-Hardebol AH, Carvalho B, Tijssen M, Belien JA, de Wit M, Delis-van Diemen PM, et al. TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression. *Gut*. 2012, 61, 1568-75.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J. STRING v10: protein protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015, 43, D447-D52.

Tanner MM, Tirkkonen M, Kallioniemi A, Holli K, Collins C, Kowbel D, et al. Amplification of Chromosomal Region 20q13 in Invasive Breast Cancer: Prognostic Implications. *Clin Cancer Res*. 1995, 1, 1455-61.

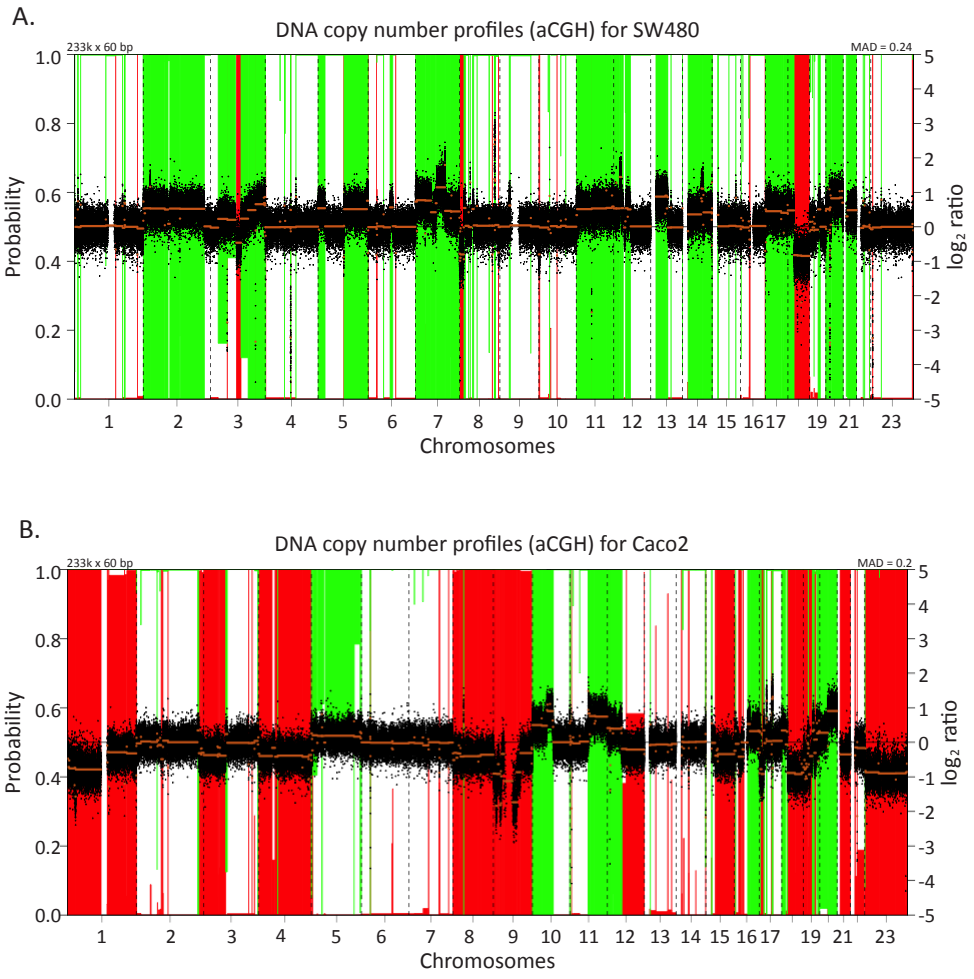
The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012, 487, 330-7.

Ubink I, Elias SG, Moelans CB, Lacle MM, van Grevenstein WMU, van Diest PJ, et al. A Novel Diagnostic Tool for Selecting Patients With Mesenchymal-Type Colon Cancer Reveals Intratumor Subtype Heterogeneity. *J Natl Cancer Inst*. 2017, 109, djw303.

Umstead M, Xiong J, Qi Q, Du Y, Fu H. Aurora kinase A interacts with H-Ras and potentiates Ras MAPK signaling. *Oncotarget*. 2017, 8, 28359-72.

Xia Z, Wei P, Zhang H, Ding Z, Yang L, Huang Z, et al. AURKA governs self-renewal capacity in glioma-initiating cells via stabilization/activation of beta-catenin/Wnt signaling. *Mol Cancer Res*. 2013, 11, 1101-11.

Zeng B, Lei Y, Zhu H, Luo S, Zhuang M, Su C, et al. Aurora-A is a novel predictor of poor prognosis in patients with resected lung adenocarcinoma. *Chin J Cancer Res*. 2014, 26, 166-73.



S3.1 Figure. DNA copy number profile (aCGH) for A) SW480 and B) Caco2, respectively. Green represents regions with gain and red represents regions with loss (also represented by the positive/ negative $\log_2$ ratio). The figures show that 20q is gained in both cells lines, SW480 otherwise has many gained regions, whereas Caco2 has many lost regions.

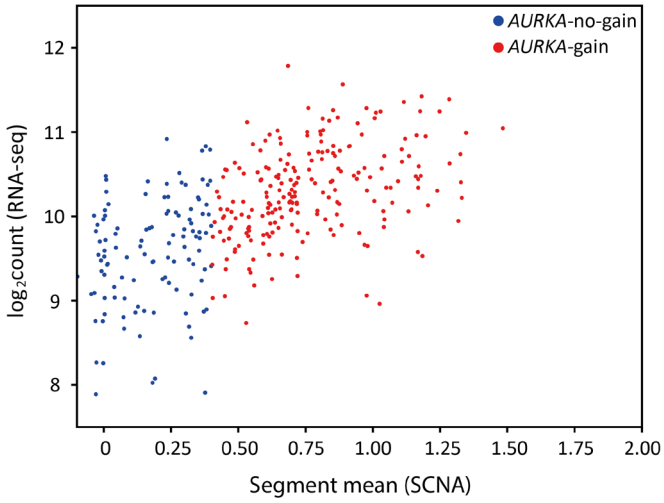
S3.1 Table. Genes in the most significantly deregulated gene modules using Heinz for SW480 with an FDR threshold of 6.21×10^{-4}.

ENSG	HUGO	*p* value	$\log_2$ FC
ENSG00000087586	AURKA	1.69×10^{-6}	-1.64
ENSG00000129625	REEP5	2.20×10^{-5}	-1.24
ENSG00000109654	TRIM2	5.38×10^{-6}	-1.16
ENSG00000213658	LAT	4.01×10^{-6}	-1.12
ENSG00000064651	SLC12A2	5.70×10^{-6}	-1.05
ENSG00000111145	ELK3	7.37×10^{-6}	-1.02
ENSG00000133703	KRAS	9.58×10^{-6}	-0.95
ENSG00000072401	UBE2D1	3.08×10^{-4}	-0.92
ENSG00000152939	MARVELD2	2.61×10^{-5}	-0.87
ENSG00000110911	SLC11A2	4.45×10^{-5}	-0.79
ENSG00000119638	NEK9	2.65×10^{-5}	-0.77
ENSG00000007168	PAFAH1B1	2.84×10^{-5}	-0.75
ENSG00000164761	TNFRSF11B	3.13×10^{-5}	-0.74
ENSG00000187109	NAP1L1	8.48×10^{-5}	-0.59
ENSG00000182866	LCK	4.12×10^{-4}	-0.49
ENSG00000177606	JUN	7.58×10^{-4}	0.47
ENSG00000145425	RPS3A	1.82×10^{-4}	0.55
ENSG00000175197	DDIT3	1.06×10^{-4}	0.61
ENSG00000169679	BUB1	3.85×10^{-5}	0.73
ENSG00000159256	MORC3	3.99×10^{-5}	0.77
ENSG00000184678	HIST2H2BE3	3.39×10^{-5}	0.78
ENSG00000185130	HIST1H2BL	2.18×10^{-5}	0.79
ENSG00000110422	HIPK3	1.41×10^{-5}	0.86
ENSG00000168036	CTNNB1	5.33×10^{-5}	0.87
ENSG00000233822	HIST1H2BN	1.09×10^{-5}	0.93
ENSG00000072571	HMMR	4.07×10^{-5}	0.95
ENSG00000277224	HIST1H2BF	2.19×10^{-5}	0.97
ENSG00000110713	NUP98	9.71×10^{-6}	0.98
ENSG00000185088	RPS27L	6.91×10^{-6}	1.05
ENSG00000152601	MBNL1	1.73×10^{-6}	1.34

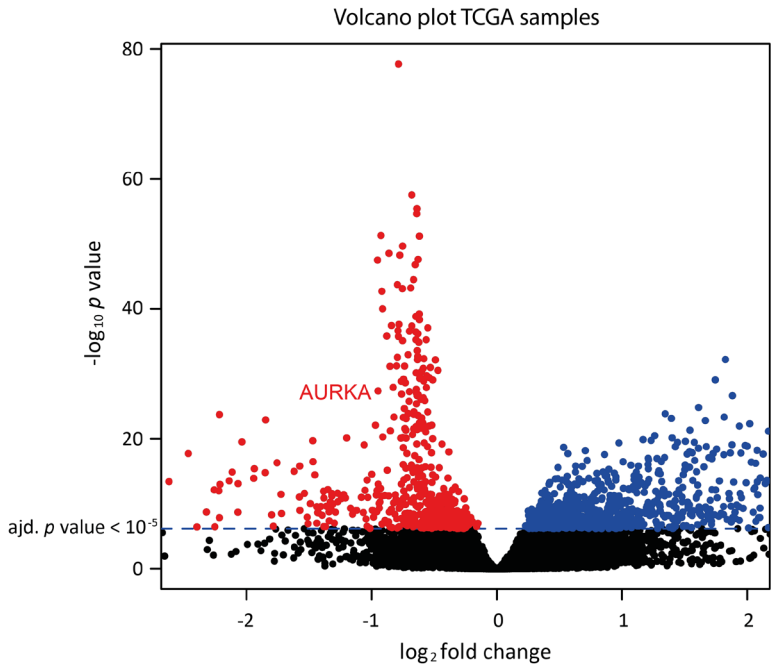
S3.2 Table. Genes in the most significantly deregulated gene modules using Heinz for Caco2 with an FDR threshold of 2.40×10^{-4}.

ENSG	HUGO	*p* value	$\log_2$ FC
ENSG00000087586	AURKA	2.28×10^{-7}	-1.31
ENSG00000164761	TNFRSF11B	3.02×10^{-7}	-1.18
ENSG00000115170	ACVR1	6.36×10^{-6}	-0.99
ENSG00000129625	REEP5	1.83×10^{-6}	-0.97
ENSG00000127022	CANX	2.32×10^{-6}	-0.95
ENSG00000179674	ARL14	1.25×10^{-6}	-0.94
ENSG00000072401	UBE2D1	1.08×10^{-5}	-0.91
ENSG00000138166	DUSP5	2.01×10^{-6}	-0.88
ENSG00000151790	TDO2	2.00×10^{-6}	-0.87
ENSG00000163631	ALB	1.01×10^{-5}	-0.85
ENSG00000143797	MBOAT2	7.00×10^{-6}	-0.84
ENSG00000154229	PRKCA	4.15×10^{-6}	-0.79
ENSG00000081051	AFP	5.31×10^{-6}	-0.76
ENSG00000116741	RGS2	1.10×10^{-5}	-0.72
ENSG00000075651	PLD1	1.26×10^{-5}	-0.71
ENSG00000106546	AHR	7.87×10^{-6}	-0.70
ENSG00000133703	KRAS	1.10×10^{-5}	-0.65
ENSG00000170345	FOS	5.35×10^{-5}	-0.51
ENSG00000168036	CTNNB1	8.65×10^{-5}	0.51
ENSG00000134853	PDGFRA	1.28×10^{-5}	0.66
ENSG00000100526	CDKN3	1.23×10^{-5}	0.66
ENSG00000188042	ARL4C	9.14×10^{-6}	0.68
ENSG00000127314	RAP1B	7.67×10^{-6}	0.70
ENSG00000204335	SP5	7.13×10^{-6}	0.71
ENSG00000110713	NUP98	6.76×10^{-6}	0.73
ENSG00000152601	MBNL1	5.91×10^{-6}	0.73
ENSG00000168646	AXIN2	6.93×10^{-6}	0.74
ENSG00000072571	HMMR	1.17×10^{-5}	0.74
ENSG00000185155	MIXL1	2.90×10^{-6}	0.85
ENSG00000105810	CDK6	4.60×10^{-6}	0.86
ENSG00000113368	LMNB1	1.85×10^{-6}	0.92

S3.2 Figure. The significantly deregulated genes from the TCGA COADREAD DEA (*AURKA*-no-gain vs. *AURKA*-gain) compared to the sets of genes significantly deregulated in the DEA of the two cell lines, SW480 and Caco2, upon *AURKA* downmodulation. $\log_2$ fold change is shown for each gene in the intersection, further illustrated by blue for upregulation and red for downregulation.



S3.3 Figure. Segment mean (SCNA; DNA copy number) plotted against $\log_2$count (RNA-seq, mRNA expression) for 330 *AURKA*-no-gain and *AURKA*-gain TCGA COADREAD samples. *AURKA*-gain samples were determined at a segment mean >= 0.4.



S3.4 Figure. Volcano plot of the genes differentially expressed between 330 *AURKA*-no-gain and *AURKA*-gain TCGA COADREAD samples. The threshold for significantly deregulated genes was set at an adjusted *p* value less than 10^{-5}.

CHAPTER 4

Differential regulation is a crucial component of the mechanism underlying inversion

bioRxiv. 2018. doi: 10.1101/449520

Saman Amini*, Annika Jacobsen*, Olga Ivanova, Philip Lijnzaad, Jaap Heringa, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren
*shared first authors

Abstract

Genetic interactions, a phenomenon whereby combinations of mutations lead to unexpected effects, reflect how cellular processes are wired and play an important role in complex genetic diseases. Understanding the molecular basis of genetic interactions is crucial for deciphering pathway organization as well as understanding the relationship between genetic variation and disease. Several putative molecular mechanisms have been linked to different genetic interaction types. However, differences in genetic interaction patterns and their underlying mechanisms have not yet been compared systematically between different functional gene classes.

Here, differences in the occurrence and types of genetic interactions are compared for two classes, gene-specific transcription factors (GSTFs) and signaling genes (kinases and phosphatases). Genome-wide gene expression data for 63 single and double deletion mutants in baker's yeast reveals that the two most common genetic interaction patterns are buffering and inversion. Buffering is typically associated with redundancy and is well understood. In inversion, genes show opposite behavior in the double mutant compared to the corresponding single mutants. The underlying mechanism is poorly understood. Although both classes show buffering and inversion patterns, the prevalence of inversion is much stronger in GSTFs. To decipher potential mechanisms, a Petri Net modeling approach was employed, where genes are represented as nodes and relationships between genes as edges. This allowed over 9 million possible three and four node models to be exhaustively enumerated. The models show that a quantitative difference in interaction strength is a strict requirement for obtaining inversion. In addition, this difference is frequently accompanied with a second gene that shows buffering.

Taken together, these results provide a mechanistic explanation for inversion. Furthermore, the ability of transcription factors to differentially regulate expression of their targets provides a likely explanation why inversion is more prevalent for GSTFs compared to kinases and phosphatases.

4.1 Introduction

Understanding the relationship between genotype and phenotype of an organism is a major challenge (Badano and Katsanis, 2002; Cooper et al., 2013). One of the difficulties for unraveling genotype-phenotype relationship has been genetic interactions, when combinations of mutations lead to phenotypic effects that are unexpected based on the phenotypes of the individual mutations (Baryshnikova et al., 2013; Phillips, 2008; Wei et al., 2014). Large-scale analyses of single and double deletion mutants have revealed that genetic interactions are pervasive in many model organisms (Babu et al., 2014; Bakal et al., 2008; Costanzo et al., 2016; Horn et al., 2011; Lehner et al., 2006; Roguev et al., 2013). Recently, efforts have been initiated to investigate genetic interactions in human cell lines too, using large-scale RNA interference and Crispr-Cas9 knock downs (Billmann et al., 2016; Han et al., 2017; Shen et al., 2017; Vizeacoumar et al., 2013). Our understanding of the molecular mechanisms that underlie genetic interactions lags behind our ability to detect genetic interactions. Understanding the molecular basis of genetic interactions and their interplay with cellular processes is important for unraveling how different processes are connected (Lehner, 2011; Sameith et al., 2015; van Wageningen et al., 2010), to what degree genetic interactions shape pathway architecture (Costanzo et al., 2016), as well as for understanding the role genetic interactions play in human disease (Moore and Williams, 2009; Wei et al., 2014).

One of the phenotypes that is frequently used to investigate genetic interactions is cell growth (Bandyopadhyay et al., 2010; Costanzo et al., 2016; Davierwala et al., 2005; Fiedler et al., 2009; Jasnos and Korona, 2007; Pan et al., 2006; St Onge et al., 2007; Szappanos et al., 2011; Tong et al., 2004; Zheng et al., 2010). Based on this phenotype, genetic interactions can be broadly subdivided in two types, negative genetic interactions where the double mutant is growing slower than expected given the growth rate of the single deletion mutants, and positive genetic interactions where the double mutant is growing faster than expected (Baryshnikova et al., 2013). Negative genetic interactions have frequently been associated with a redundancy relationship between two functionally related genes (Hartman et al., 2001). The redundancy mechanisms by which two genes can compensate for each other's loss has been linked with close paralog genes or redundant pathways (Amini et al., 2017; Ihmels et al., 2007). Positive genetic interactions have been associated with genes participating in the same protein complex or pathway (Collins et al., 2007). There are however many exceptions to these rules and it also has become clear that there are many other potential mechanisms underlying these genetic interactions (Baryshnikova et al., 2013; Lehner, 2011).

Another phenotype that has been less frequently used for investigating genetic interactions is gene expression (Capaldi et al., 2008; Dixit et al., 2016; Gutin et al., 2015; Pirkl et al., 2017; Sameith et al., 2015; van Wageningen et al., 2010). Expression-based genetic interaction profiling provides detailed information at the molecular level, which is beneficial for unraveling mechanisms of genetic interactions (Capaldi et al., 2008; Dixit et al., 2016; Gutin et al., 2015; Pirkl et al., 2017; Sameith

et al., 2015; van Wageningen et al., 2010). Unlike growth-based profiling, which gives a subdivision into either positive or negative interactions, expression-based genetic interaction profiling provides further subdivision into more specific genetic interaction patterns including buffering, quantitative buffering, suppression, quantitative suppression, masking and inversion (Sameith et al., 2015). A more detailed sub classification that includes information on expression of downstream genes, can also contribute to understanding the mechanisms by which two genes interact (Sameith et al., 2015; van Wageningen et al., 2010; Wong et al., 2016).

To provide mechanistic insights into biological networks, Boolean modeling has been used successfully (Kauffman et al., 2003; Li et al., 2004). It has also been applied to unravel regulatory networks underlying genetic interaction patterns between kinases and phosphatases (van Wageningen et al., 2010). Due to their intrinsically simple nature, such Boolean network models allow exhaustive enumeration of network topologies. The outcomes of these models can then be easily compared to the patterns observed in experimental data. Boolean operators however, are limited to on and off values and cannot easily accommodate quantitative measurements, which limits the types of genetic interaction patterns that can be investigated using this approach. Unraveling the regulatory network underlying genetic interaction patterns would potentially benefit from application of modeling approaches that allow some degree of quantitativeness to be introduced while still being computationally feasible to exhaustively explore all potential models. In this way, Petri nets may be considered an extension of Boolean modeling that provides more flexibility, in particular by choosing different network edge strengths, without the need to incorporate detailed prior quantitative knowledge (Bonzanni et al., 2014; Bonzanni et al., 2013; Bonzanni et al., 2009; Chaouiya et al., 2006). Petri net modeling would therefore allow investigation of all possible genetic interaction patterns in an exhaustive and semi-quantitative manner.

It is evident that genetic interactions are widespread in *Saccharomyces cerevisiae* (Costanzo et al., 2016) as well as other organisms (Babu et al., 2014; Lehner et al., 2006). Nevertheless, extensive characterization of the molecular mechanisms underlying genetic interactions, as well as a comparison of the molecular mechanisms underlying genetic interactions between different functional classes have, as yet, not been performed. Here, two functional classes, gene specific transcription factors (GSTFs) and signaling related genes (kinases and phosphatases) have been compared with regard to negative genetic interaction patterns and the possible underlying molecular mechanisms. This revealed that the two most common genetic interaction patterns are buffering and inversion. The prevalence of inversion however, is much stronger in GSTFs. The underlying mechanism of inversion, whereby genes show opposite behavior in the double mutant compared to the corresponding single mutants, is poorly understood. Exhaustive enumeration of network topologies using Petri net modeling reveals that the minimum requirement for observing inversion is having a quantitative difference in interaction strength (edge weight) from the two upstream transcription factors to a shared downstream gene. In addition, this quantitative edge difference is frequently accompanied

by an intermediate node, that displays a buffering pattern. The proposed model provides a mechanistic explanation for inversion, thereby further aiding a better understanding of genetic interactions. GSTFs, more so than kinases/phosphatases, can modulate or fine-tune the activation levels of their target genes, which suggests quantitative differences in regulating downstream target genes are important for the functioning of GSTFs. This is consistent with the fact that inversion occurs more often between GSTFs than between signaling genes, as well as our observation that quantitative edge differences are required for inversion to occur and provides a likely explanation why inversion is more prevalent for transcription factors.

4.2 Results

4.2.1 A single dataset to compare mechanisms of genetic interactions between gene specific transcription factors and kinases/phosphatases

To investigate potential differences in mechanisms of genetic interactions between groups of genes with a different function, data from two previously published datasets were combined (Sameith et al., 2015; van Wageningen et al., 2010). The first dataset includes genome-wide gene expression measurements of 154 single and double gene-specific transcription factor (GSTF) deletion mutants (Sameith et al., 2015). The second dataset contains genome-wide gene expression measurements of 54 single and double kinase/phosphatase (K/P) deletion mutants (van Wageningen et al., 2010). These studies applied different criteria to select for interacting pairs. Whereas the GSTF dataset includes both positive and negative genetic interactions, the kinase/phosphatase dataset was restricted to negative genetic interactions only. To avoid potential biases, the selection criteria of the kinase/phosphatase dataset (van Wageningen et al., 2010) were adopted and applied to both datasets. In short, selection was based on pairs having a significant growth-based negative genetic interaction score ($p < 0.05$, Methods) to include redundancy relationships that influence fitness. In addition, for a given double mutant, at least one of the corresponding single mutants has an expression profile similar to wildtype (WT) (eight or more transcripts changing significantly ($p < 0.05$, fold-change > 1.7)) to ensure that genetic interactions such as redundancy are considered. These selection criteria yield a uniform dataset consisting of 11 GSTF double mutants and 15 kinase/phosphatase double mutants as well as their respective single mutants (63 single and double mutants in total; S4.1 Table).

4.2.2 Genetic interaction profiles indicate a large degree of buffering

Genetic interactions can be investigated in different ways. Here, both growth as well as genome-wide gene expression is used to compare genetic interactions between GSTFs and kinases/phosphatases, as described before (Sameith et al., 2015). In short, a growth-based genetic interaction score $\varepsilon_{growth,XY}$ between two genes X and Y is obtained by comparing the observed fitness for double mutant $W_{x\Delta y\Delta}$ to the fitness that is expected based on both single mutants $W_{x\Delta} \cdot W_{y\Delta}$ ($\varepsilon_{growth,XY} = W_{x\Delta y\Delta} - W_{x\Delta} \cdot W_{y\Delta}$)

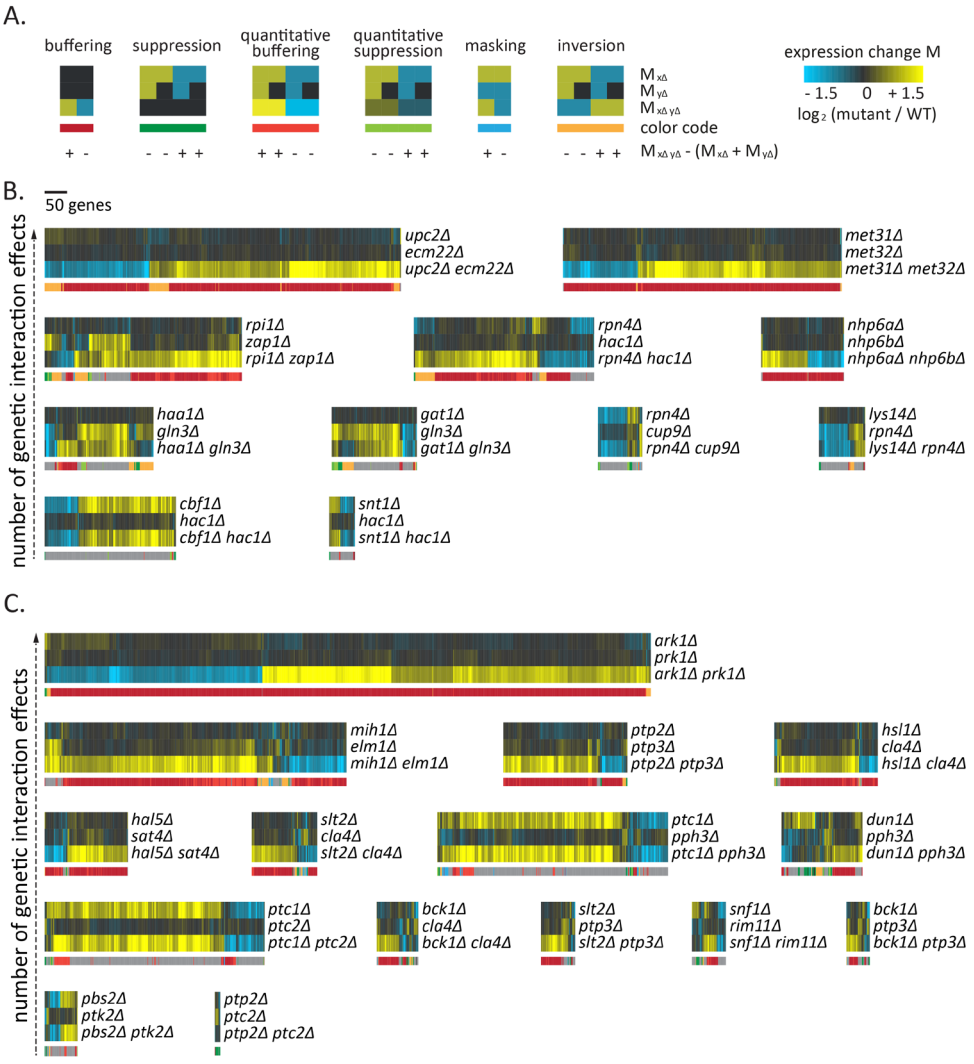


Figure 4.1. Genetic interaction profiles of GSTF and kinase/962 phosphatase pairs. (A) Cartoon depicting expression changes in single and double mutants with different genetic interaction patterns color coded underneath. At the bottom, the direction of expression differences between the observed expression change ($M_{xΔyΔ}$) and expected ($M_{xΔ}$+$M_{yΔ}$) is stated. Color scale from yellow for an increase in expression levels compared to WT ($p \leq 0.01$, $log_2(FC) > 0$), black for unchanged expression ($p > 0.01$) and blue for a decrease in expression levels compared to WT ($p \leq 0.01$, $log_2(FC) < 0$). (B) Expression changes compared to WT (horizontal) in GSTF single and double mutants (vertical). Different colors underneath the gene expression profiles represent different genetic interaction patterns as indicated in A. Gray depicts gene expression changes not part of a genetic interaction pattern. Pairs are sorted based on the number of genetic interaction effects, increasing from bottom to top. (C) Expression changes compared to WT (horizontal) in kinase and phosphatase single and double mutants (vertical). Layout and ordering as in B.

(Mani et al., 2008). A gene expression-based genetic interaction score between two genes X and Y is calculated in two consecutive steps (Sameith et al., 2015). First, the effect of a genetic interaction between two genes X and Y on any downstream gene *i* is calculated as the deviation between the expression change observed in the double mutant $M_{i,x\Delta y\Delta}$ and the expected expression change based on the corresponding single mutants $M_{i,x\Delta} + M_{i,y\Delta}$ ($\varepsilon_{txpm\_i,XY} = |M_{i,x\Delta y\Delta} - (M_{i,x\Delta} + M_{i,y\Delta})|$). The overall genetic interaction score between gene X and Y is then obtained by counting the total number of genes for which $\varepsilon_{txpm\_i,XY}$ is greater than 1.5 (Sameith et al., 2015). Gene expression changes from single and double mutants were subsequently grouped into the six genetic interaction patterns, buffering, suppression, quantitative buffering, quantitative suppression, masking and inversion, as previously described (Figure 4.1A) (Sameith et al., 2015). When investigating the genetic interaction profiles of GSTFs (Figure 4.1B) as well as kinases/phosphatases (Figure 4.1C), it is clear that buffering is prevalent in many of the larger genetic interaction profiles, but the degree of buffering differs for the smaller genetic interaction profiles.

4.2.3 Removal of a slow growth associated expression signature for improved identification of direct effects

Hierarchical clustering was applied to group pairs with similar genetic interaction patterns (S4.1 Figure), thereby disregarding the identity of individual downstream genes. From this clustering, it is clear that there is no distinct separation between pairs consisting of GSTFs and kinases/phosphatases. Instead, most pairs are characterized by large buffering effects, grouped together in a single large cluster (S4.1A Figure, red branch labeled as 1). This is not surprising, since all pairs are selected for having a significant growth-based negative genetic interaction score. This in turn is based on double mutants growing slower than expected based on the single mutants. Slow growing strains are known to display a common gene expression signature (Keren et al., 2013; Regenberg et al., 2006). This slow growth gene expression signature is caused by a change in the distribution of cells over different cell cycle phases (O'Duibhir et al., 2014). To facilitate investigating mechanisms of genetic interactions, such effects are better disregarded. As described previously (O'Duibhir et al., 2014), the dataset was transformed by removing the slow growth signature (Methods). Removing the slow growth signature and thereby reducing effects due to a cell cycle population shift improves identification of direct target genes of GSTF pairs (S4.2 Figure) as shown before for individual GSTFs (O'Duibhir et al., 2014).

4.2.4 Discerning potential mechanisms with slow growth corrected genetic interaction profiles

Hierarchical clustering of the slow growth corrected genetic interaction profiles was then applied to unravel potential differences in observed genetic interactions patterns between GSTFs and K/P (Figure 4.2A-C). Three striking differences emerge when comparing this clustering with the clustering of the original, untransformed data (S4.1 Figure). First, pairs are grouped into four distinct clusters,

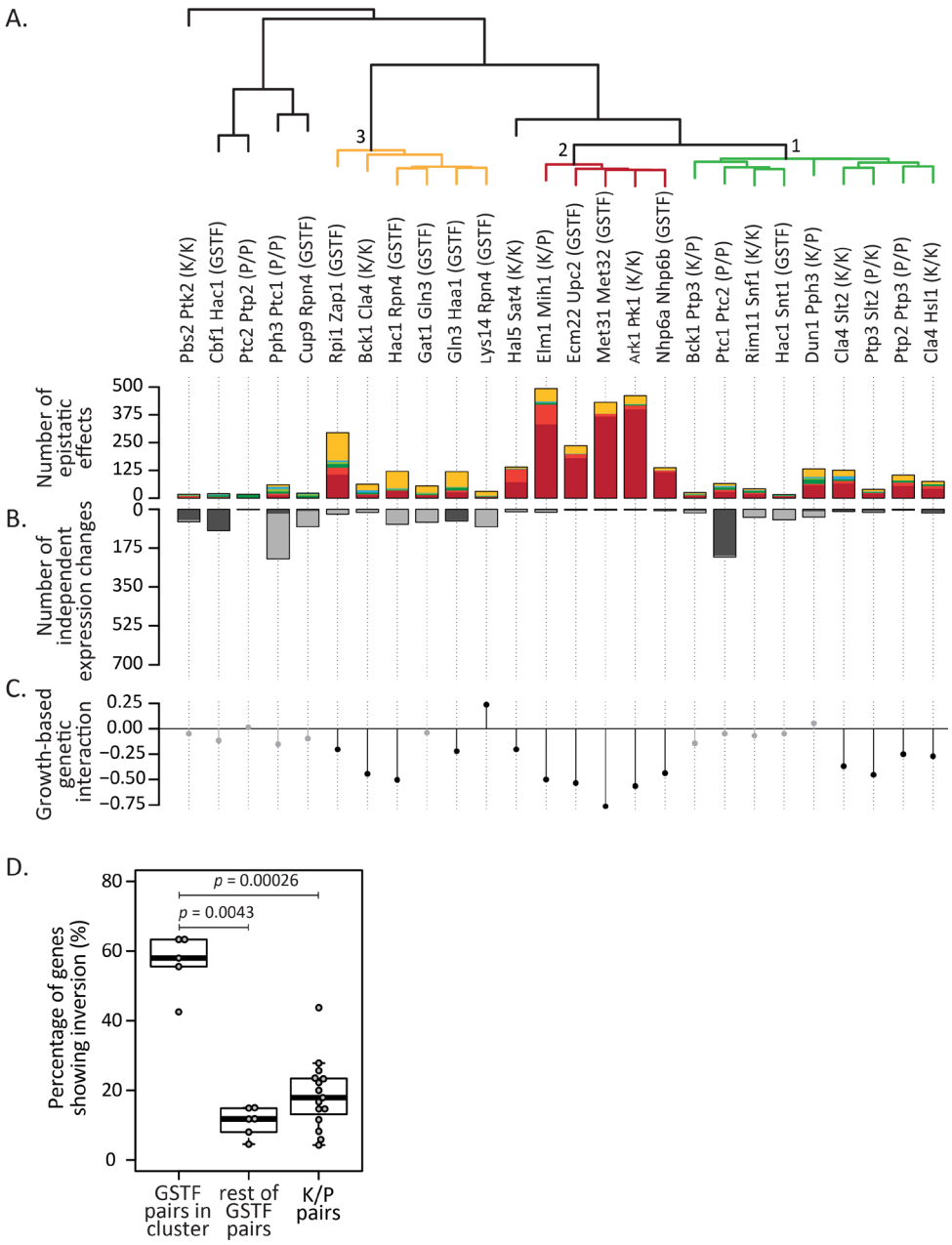


Figure 4.2. Hierarchical clustering of slow growth corrected genetic interaction profiles is better suited to discern underlying mechanisms. A) Hierarchical clustering of all pairs according to their genetic interaction effects after slow growth correction. Average linkage clustering was applied to group pairs with similar genetic interaction patterns. The number of occurrences for each genetic interaction pattern (Fig 1A) was used and the identity of individual genes was disregarded. Similarity between pairs was calculated using cosine correlation. Branch depicted in red, label 2, indicates pairs that are dominated by buffering. Branch depicted in orange, label 3, indicates pairs dominated by

(Continuing caption for Figure 4.2 on previous page).

inversion. Branch depicted in green, label 1, indicates pairs explained by mixed epistasis. The number of genetic interaction effects underlying the clustering are shown as bar plots below the dendrogram (colors as in Fig 1A). B) Number of genes showing no genetic interaction pattern but significantly changing in one of the mutants compared to WT ($p \leq 0.01$, FC > 1.5). Dark gray for the first named gene, light gray for the second named gene. C) Growth-based genetic interaction scores depicted by solid circles. Significant genetic interaction scores are shown in black, gray otherwise. Ordering of pairs is the same as in A and B. D) Boxplot highlighting the difference between the percentage of genes showing inversion for GSTF pairs within the orange branch (Fig 2A), GSTF pairs outside this cluster and K/P pairs. *p* values are based on a two-sided Mann-Whitney test.

whereas previously, most were grouped into a single large cluster. Second, a cluster of predominantly kinase/phosphatase pairs emerges (Figure 4.2A, green branch, labeled as 1). These contain mixtures of different genetic interaction patterns, corresponding to 'mixed epistasis' (van Wageningen et al., 2010). Third, a smaller cluster dominated by buffering appears (Figure 4.2A, red branch, labeled as 2). This cluster also has strong growth-based negative genetic interaction scores (Figure 4.2C), which is known to be associated with redundancy. The 'buffering' cluster, with its strong growth-based negative interactions, mostly consists of pairs with a high sequence identity (average 43.7%) compared to the others (average 21%). These include Nhp6a-Nhp6b, Met31-Met32, Ecm22-Upc2 and Ark1-Prk1, for all of which redundancy relationships have been described previously (Blaiseau et al., 1997; Cope et al., 1999; Costigan et al., 1994; Vik and Rine, 2001). The high sequence identity here indicates a homology-based redundancy, in which both genes can perform the same function (Amini et al., 2017; Ihmels et al., 2007; Keane et al., 2014; Plata and Vitkup, 2014). The only exception here is the kinase/phosphatase pair Elm1-Mih1. This pair may be explained through pathway-based redundancy where two parallel pathways can compensate for each other's function (Boone et al., 2007). Elm1 is a serine/threonine kinase, and Mih1 a tyrosine phosphatase, which are both involved in cell cycle control (S4.3 Figure, left panel) (Bouquin et al., 2000; Russell et al., 1989). Mih1 directly regulates the cyclin-dependent kinase Cdc28, a master regulator of the G2/M transition (Russell et al., 1989). Elm1, on the other hand, indirectly regulates Cdc28 activity by promoting Swe1 degradation through the recruitment of Hsl1 (McMillan et al., 1999; Thomas et al., 2003). The timing of entry into mitosis is controlled by balancing the opposing activities of Swe1 and Mih1 on Cdc28, and both Swe1 and Mih1 are key in the checkpoint mediated G2 arrest (Morgan, 1997; Russell, 1998). Deletion of Elm1 does not result in many gene expression changes (Figure 4.1C), which can be explained through compensatory activity of Mih1 (S4.3 Figure, middle panel). Downregulation of Mih1 activity has also been suggested before as an effective mechanism to counter stabilization of Swe1, as neither stabilization of Swe1 or elimination of Mih1 in itself is sufficient to promote G2 delay, but simultaneous stabilization of Swe1 and elimination of Mih1 does cause G2 arrest (McMillan et al., 1999). Simultaneous deletion of Elm1 and Mih1 leads to higher levels of inactive Cdc28 causing a G2 delay and stress (S4.3 Figure, right panel) (McMillan et al., 1999). All pairs within this cluster can

therefore be associated with a redundancy mechanism. Taken together, these results suggest that the clustering of the slow growth corrected genetic interaction profiles is able to discern potential differences in mechanisms. Even though most pairs in the four clusters (Figure 4.2A) show negative genetic interactions (Figure 4.2C), different mechanisms are likely underlying each individual cluster.

4.2.5 Inversion is associated with a specific subset of GSTFs

Within the slow growth corrected genetic interaction profiles another interesting cluster stands out: the orange branch where five out of six pairs involve GSTFs, which predominantly show the inversion pattern (Figure 4.2A, branch 3). This suggests that inversion may be strongly associated with a particular group of GSTFs, whereas this does not seem to be the case for kinases and phosphatases. The overall percentage of genes showing inversion is already much higher for GSTFs (28.6%) than for kinases/phosphatases (18.7%) (S4.2 Table). When investigating the GSTF pairs within the cluster, it is clear that these display an even higher percentage of inversion compared to kinases and phosphatases (Figure 4.2D; $p = 0.00026$) as well as compared to other GSTF pairs (Figure 4.2D; $p = 0.0043$). In order to determine whether inversion was specific to the set of GSTFs analyzed here, or part of a more general phenomenon common to GSTFs, we included both positive and negative genetic interactions between GSTF pairs, expanding the number of GSTF pairs to 44. Clustering of all 44 GSTF pairs (S4.4 Figure) also shows that a large fraction of the GSTF pairs contain many genes showing inversion, with most of the inversion dominated GSTF pairs still clustering together (S4.4 Figure, indicated with an asterisk). Note though, that because the 44 GSTF pairs include both positive and negative genetic interactions, the results are not directly comparable to the kinase/phosphatase pairs as these only include negative genetic interactions. Taken together, this indicates that not only is inversion more frequently associated with GSTFs compared to kinases and phosphatases, but one particular subset of GSTFs is also predominantly defined by inversion.

4.2.6 An exhaustive modeling approach to explore potential mechanisms underlying inversion

Unlike buffering, where redundancy is a likely mechanistic explanation, the underlying mechanism of inversion is still unknown (Sameith et al., 2015). The GSTF pairs within the inversion dominated cluster also do not share a common biological process, function, pathway or protein domain other than general transcription related processes and functions. To investigate potential mechanisms of inversion, an exhaustive exploration was initiated. Previously, Boolean modeling has been applied to exhaustively explore all mechanisms underlying two genetic interaction patterns for the Fus3-Kss1 kinase phosphatase pair (van Wageningen et al., 2010). However, to explore all potential mechanisms underlying inversion, a Boolean approach may not suffice as more subtle, quantitative effects, may be needed to obtain inversion. At the same time, any modeling approach must remain computationally feasible. For this purpose, a modeling approach based on Petri

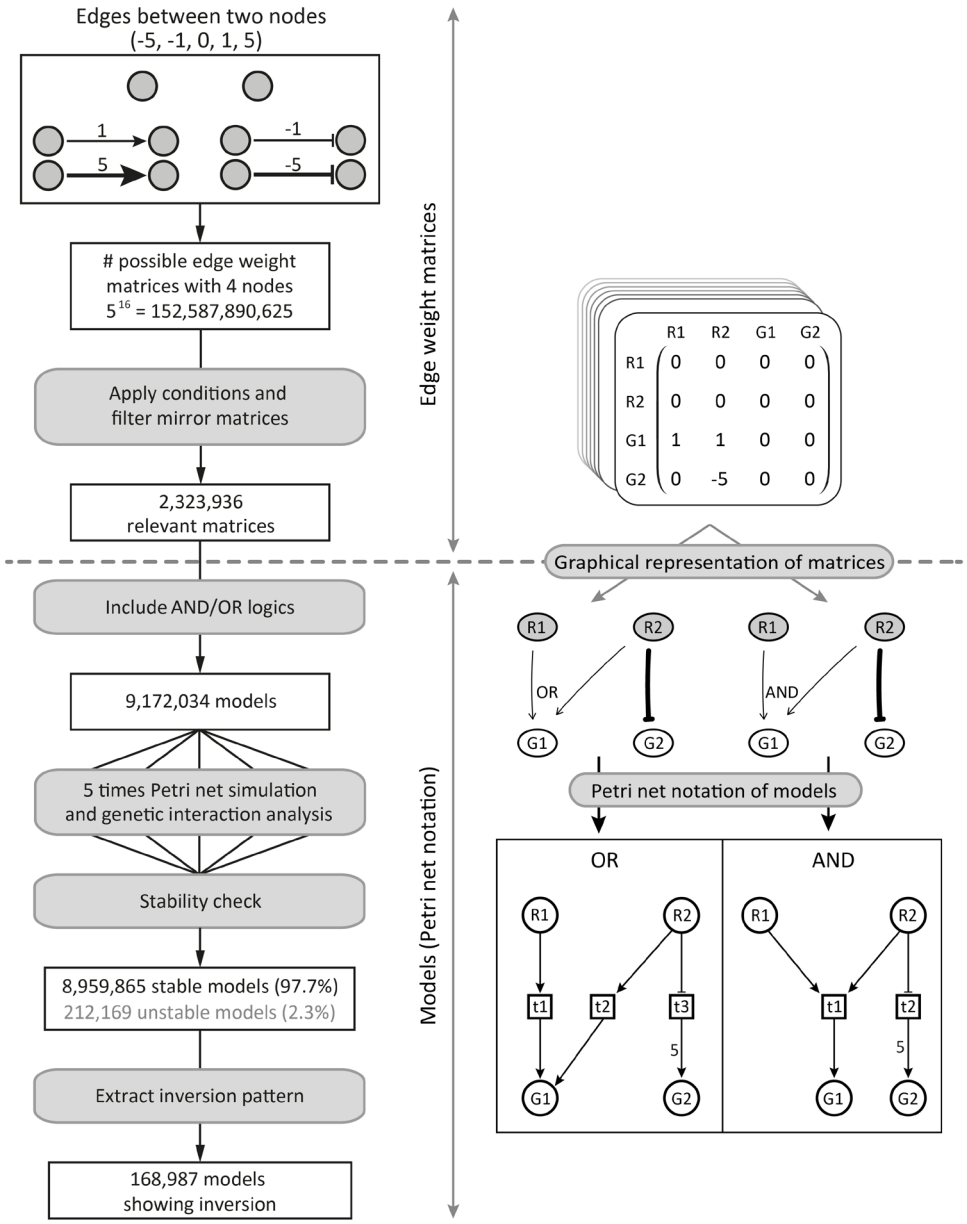


Figure 4.3. Schematic overview of Petri net simulation pipeline. Schematic overview of the pipeline implemented for performing Petri net simulations. The left panels show from top to bottom the different steps performed when running the simulation pipeline. The right panels show the different data representations used throughout the pipeline. The right panel above the dashed line indicates a series of steps where edge weight matrices are used. The right panel below the dashed line indicates steps where models or Petri net notation are used.

nets was devised to exhaustively evaluate all possible three and four node models but taking into account the possibility of quantitatively different effects (Figure 4.3, Methods). Interactions between nodes (edges) can be activating (positive) or inhibiting (negative). In order to incorporate quantitative differences, both strong and weak edges were used (Methods). Counting all possible combinations of different edges results in 152,587,890,625 possible edge weight matrices. To reduce the number of models, three conditions were imposed, as used previously (van Wageningen et al., 2010). In short, nodes contain no self-edges, the number of incoming edges on any node is limited to two and the model includes at least two edges from one of the regulators (R1, R2) to the downstream genes (G1, G2). Applying these requirements and filtering for mirror edge weight matrices results in 2,323,936 matrices. By including AND/OR logics the final number of models to be evaluated was 9,172,034 (Methods). Petri net simulations were then run and genetic interaction patterns determined for G1 and G2, analogous to what was done for the original data (Methods) (Figure 4.1A). Depending on the topology, Petri net models can be stochastic, in other words, they do not show the same behavior when simulated multiple times and therefore result in unstable models. Only 2.3% of the models were found to be unstable, i.e. showed inconsistent genetic interaction patterns for G1 and G2 across five times simulation runs. Thus, stochasticity hardly influences the observation of genetic interaction patterns in our simulations (Figure 4.3). Nevertheless, unstable models were excluded from further analysis. In total, 168,987 models (1.8%) show inversion in either G1, G2, or both downstream nodes.

4.2.7 A quantitative difference in interaction strength is a strict requirement when observing inversion

To investigate which potential regulatory patterns underlie the 168,987 models showing inversion, low complexity models with few edges were analyzed first. Two interesting observations can be made. First, although there are many high complexity models involving four nodes and many edges (up to eight), three nodes and three edges are sufficient to explain inversion (Figure 4.4A). Second, only two three-node models exist that show inversion (Figure 4.4A). These two models only differ in the strength of the inhibiting edge from R1 to R2. Both models involve inhibition of R2 through R1 and weak activation of G1 by R1 in combination with a strong activation of G1 by R2, i.e. a quantitative edge difference between the incoming edges of G1. Deletion of R1 in these two models results in activation of R2, and therefore upregulation of G1 due to a strong activating edge. Deletion of R2 however, will not result in any changes compared to WT as it is normally inhibited by R1. Deletion of both R1 and R2 will lead to downregulation of G1 as the weak activating edge from R1 to G1 is lost. Taken together, the analysis of the low complexity models indicates that a quantitative difference in interaction strength is required to explain inversion. To investigate whether this requirement also holds for higher complexity models, all models containing two to eight edges were further analyzed. Inversion models were grouped by the number of edges

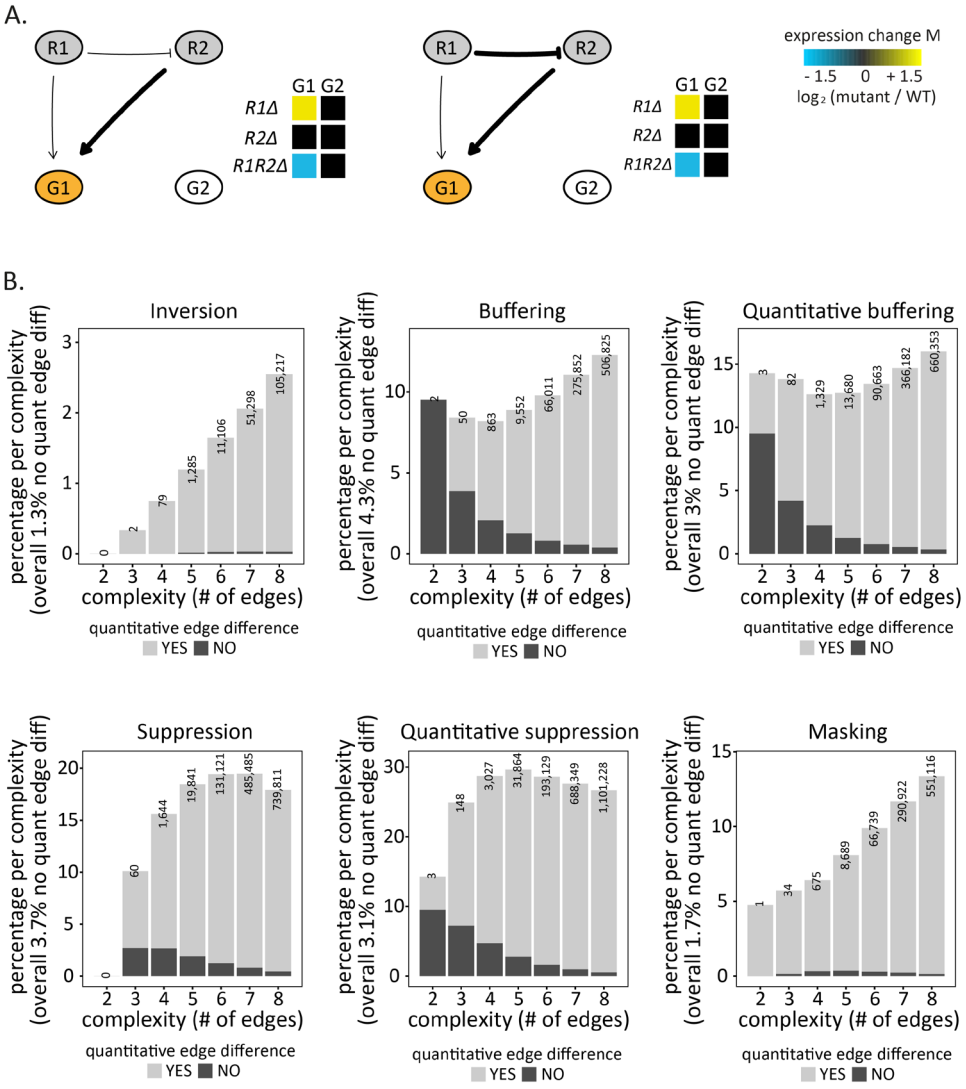


Figure 4.4. A quantitative edge difference is the minimum requirement for observing inversion. A) Petri net simulation results for the only two models with three nodes that result in inversion (indicated in orange) for the G1 node. Heat maps indicate the $\log_2$(FC) of the number of tokens in simulated deletion mutants (single and double mutant) relative to the WT situation. Thicker lines indicate edges with a strong effect. B) For each genetic interaction pattern (inversion, buffering, quantitative buffering, suppression, quantitative suppression and masking), the percentage of models showing that particular genetic interaction pattern is shown, split up per complexity (number of edges). The percentage per complexity is calculated as the number of models showing a particular genetic interaction pattern for a certain complexity, divided by the total number of models for that complexity. Bar plots are subdivided into two types of models, models that have quantitative differences between edge weights (bright gray) and models that have no quantitative differences between edge weights (dark gray). The number of models showing the particular genetic interaction pattern per complexity is shown on top of each bar plot.

(complexity) and then analyzed for their relative frequency of having a quantitative edge difference (Figure 4.4B, top left panel, note that the number of possible models grows exponentially with the number of edges). Almost all of these models show a quantitative edge difference, with only a very small fraction (1.3% overall) of models not having a quantitative edge difference. Except for masking, the other genetic interaction patterns show different behavior, indicating that the relative ratio of quantitative versus non-quantitative edges is not an inherent network property. Based on both the low complexity models as well as the high complexity models showing inversion, it is evident that a quantitative difference in interaction strength of two genes or pathways acting on a downstream gene is required to explain inversion.

4.2.8 A quantitative difference in interaction strength is frequently accompanied by an intermediate buffering node

With the exception of the two models discussed above, all other inversion models consist of four nodes with two regulator nodes and two downstream effector nodes. To better understand the interplay between all four nodes, besides the node

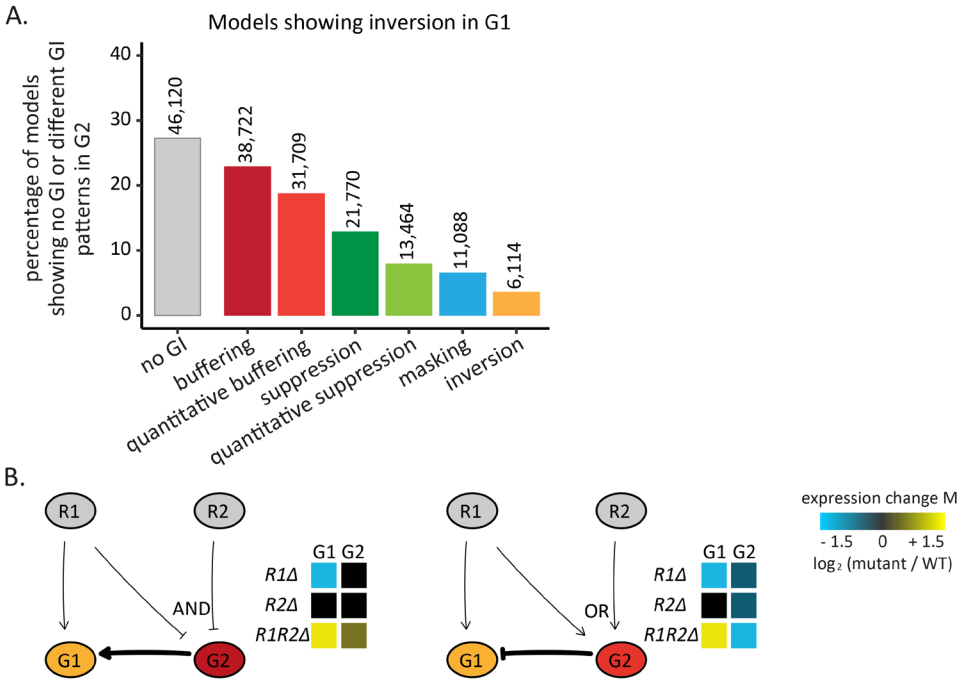


Figure 4.5. Inversion is frequently accompanied by buffering. A) Bar plots showing the percentage of models that either have no genetic interaction (gray, left bar) or a different genetic interaction pattern in node G2 when node G1 is displaying inversion. The number of models per category is shown on top of each bar plot. Color scheme of the genetic interaction patterns as in Fig 1A. B) Petri net simulation results for two models with four nodes with node G1 always displaying inversion and node G2 displaying either buffering (left) or quantitative buffering (right). Heat maps as in Fig 4A.

Table 4.1. Models with a quantitative edge difference and intermediate buffering node.

		Quantitative edge difference	
		YES	NO
Buffering or quantitative buffering	YES	69,333 (41.03%) * 1,754,000 (23.66%) #	1,098 (0.65%) * 65,618 (0.89%) #
	NO	97,418 (57.65%) * 5,412,614 (73.01%) #	1,138 (0.67%) * 180,808 (2.44%) #

Inversion models are indicated with *. All models are indicated with #. The combination of a quantitative difference in edge strength and buffering is enriched for the inversion models (41% vs. 24%, $p < 0.005$).

displaying inversion (G1), the second downstream gene (G2) was also analyzed for the occurrence of different genetic interaction patterns (Figure 4.5A). Most G2 nodes tend to have no genetic interaction pattern (27%). The most common genetic interaction patterns are buffering (23%) and quantitative buffering (18%). These both are very alike in their genetic interaction pattern (Figure 4.1A) and only show slight differences in their quantitative behavior. They may therefore be considered as part of the same superclass of "buffering". The buffering node is frequently positioned upstream of the inversion node, and always downstream of R1/R2 (Figure 4.5B). The combination of inversion and buffering is also significantly overrepresented within inversion models when compared to all models (Table 4.1, $p < 0.005$). Taken together this shows that a quantitative difference in interaction strength of two genes or pathways acting on a downstream gene is frequently accompanied by an intermediate gene or pathway that displays buffering.

4.2.9 Gat1 and Gln3 might differentially regulate mitochondrial-to-nuclear signaling

One gene pair within the inversion dominated GSTF cluster (Figure 4.2A, branch 3; Figure 4.6A) that largely consists of inversion is Gat1-Gln3. By combining the three-node model derived from the Petri Net modeling (Figure 4.4A, left panel) with existing literature, a potential mechanistic explanation for the interaction between this pair can be obtained (Figure 4.6B). Both Gln3 and Gat1 are activators involved in regulating nitrogen catabolite repression (NCR) sensitive genes (Coffman et al., 1996; Minehart and Magasanik, 1991; Stanbrough et al., 1995). When cells are grown under nitrogen rich conditions, as was done here, Gat1 is repressed by Dal80 (Coffman et al., 1996). Dal80 in turn can be activated by Gln3 (Coffman et al., 1996; Cunningham and Cooper, 1991), which provides a plausible mechanism for the predicted inhibition edge between Gln3 and Gat1 (Figure 4.6B). The degree to which Gln3 and Gat1 influence downstream genes has also been reported to differentiate between individual genes (Saxena et al., 2003), which is fully consistent with the quantitative edge difference as predicted in the model (Figure 4.6B). The set of inversion related genes (Figure 4.6A, gene set 1) is enriched for nuclear encoded mitochondrial respiratory genes (Figure 4.6A, denoted with a dot, p value 3.20×10^{-17}). Previously, NCR has been linked with mitochondrial-to-nuclear

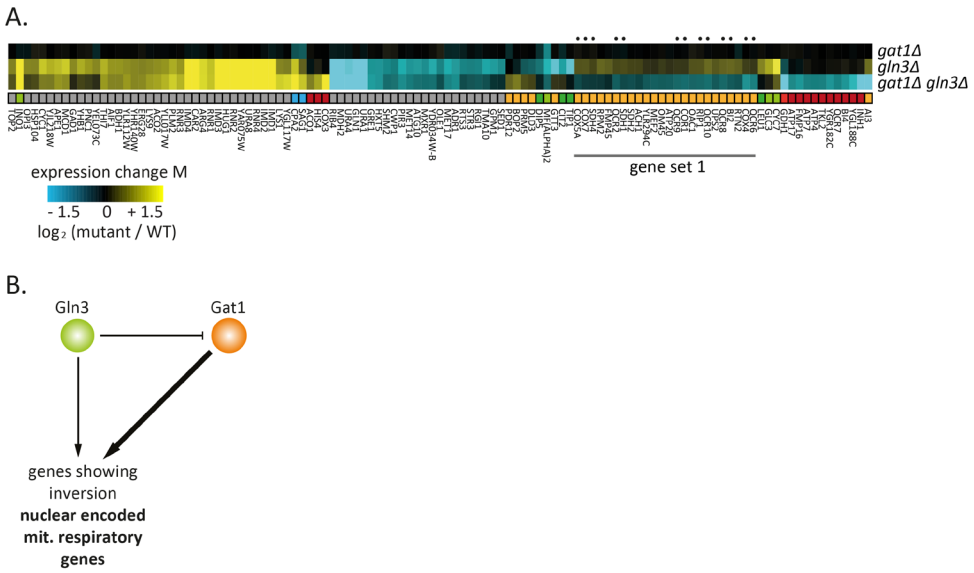


Figure 4.6. Gln3 and Gat1 might differentially regulate mitochondrial to nuclear signaling. A) Expression changes compared to WT (horizontal) in *gat1Δ*, *gln3Δ*, and *gat1Δ gln3Δ* mutants (vertical) after slow growth correction. Different colors underneath the gene expression profiles represent different genetic interaction patterns as indicated in Fig 1A. Gray depicts gene expression changes not part of a genetic interaction pattern. Nuclear encoded mitochondrial respiratory genes are denoted with a dot. B) Proposed model to explain the inversion pattern between Gat1 and Gln3 based on the Petri net simulation result in Fig 4A.

signaling through the retrograde signaling pathway (Butow and Avadhani, 2004; Giannattasio et al., 2005), although an alternative mitochondrial-to-nuclear signaling pathway, such as the intergenomic signaling pathway, may instead be involved (Dagsgaard et al., 2001). Taken together, this suggests that Gat1 and Gln3 might differentially influence mitochondrial-to-nuclear signaling, although additional experiments would be needed to confirm this initial hypothesis.

4.2.10 Pdr3 likely acts as the intermediate buffering gene in mediating the inversion pattern observed for Hac1-Rpn4

Another interesting pair of genes within the GSTF cluster dominated by the inversion pattern (Figure 4.2A, branch 3) is Hac1-Rpn4. This pair displays a substantial amount of both inversion as well as buffering (Figure 4.7A) and lends itself well for testing some of the model predictions. Hac1 and Rpn4 are both involved in the processing of inappropriately folding proteins, either by activating genes of the unfolded protein response (Mori et al., 1996) (UPR, Hac1) or via the endoplasmic reticulum-associated degradation (Mannhaupt et al., 1999) (ERAD, Rpn4). Two genes that display inversion, Pdr5 and Pdr15, show stronger expression changes compared to the other genes in the same gene set (Figure 4.7A, gene set 1). Both Pdr5 and Pdr15 are multidrug transporters involved in the pleiotropic drug

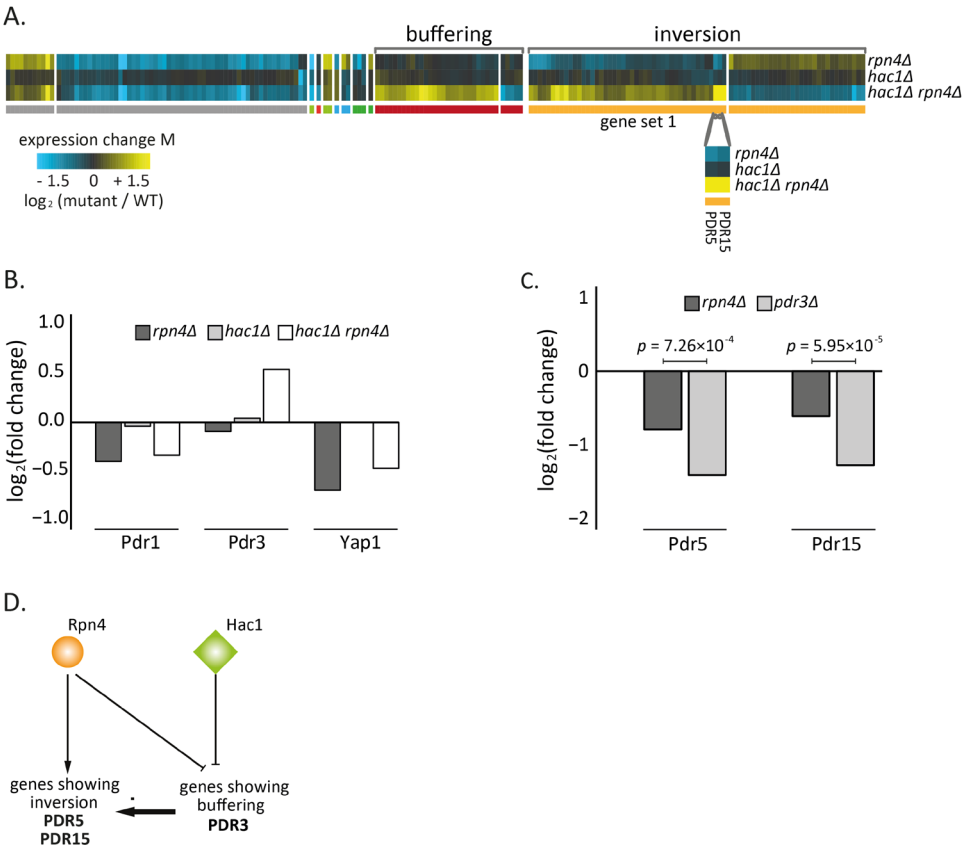


Figure 4.7. Pdr3 acts as an intermediate gene for observing inversion in PDR5 and PDR15. A) Expression changes compared to WT (horizontal) in *rpn4Δ*, *hac1Δ*, and *hac1Δ rpn4Δ* mutants (vertical) after slow growth correction. Different colors underneath the gene expression profiles represent different genetic interaction patterns as indicated in Fig 1A. Gray depicts gene expression changes not part of a genetic interaction pattern. B) Expression changes of Pdr1, Pdr3 and Yap1 compared to WT in *rpn4Δ*, *hac1Δ* and *hac1Δ rpn4Δ* mutants. C) Expression changes of Pdr5 and Pdr15 compared to WT in *rpn4Δ* and *pdr3Δ* mutants. *p* values are obtained from a limma analysis comparing gene expression changes between *rpn4Δ* and *pdr3Δ* mutants. D) Proposed model to explain the inversion pattern between Hac1 and Rpn4 based on the Petri net simulation result in Fig 5B.

response (Golin et al., 2007). Expression of these two genes is tightly regulated by Pdr1 and Pdr3 (Katzmann et al., 1994; Wolfger et al., 1997). Pdr5 is also positively regulated by expression of Yap1, a basic leucine zipper transcription factor that is required for oxidative stress tolerance (Miyahara et al., 1996). Of the three transcription factors Pdr1, Pdr3 and Yap1, only PDR3 shows a clear upregulation in the *hac1Δ rpn4Δ* double mutant and hardly any change in the respective single mutants (Figure 4.7B). This is consistent with the role of the intermediate buffering gene as derived from our Petri net modeling results. If Pdr3 acts as the intermediate buffering gene as predicted based on our model, it is also expected that deletion

of PDR3 leads to a more severe downregulation of *PDR5* and *PDR15* expression levels when compared to expression levels of *PDR5* and *PDR15* in the *rpn4Δ* mutant. To test this prediction, mRNA expression changes of *PDR5* and *PDR15* where investigated in the *pdr3Δ* and *rpn4Δ* mutants. As expected, deletion of PDR3 results in a much stronger downregulation of *PDR5* ($p = 7.26\times10^{-4}$) and *PDR15* ($p = 5.95\times10^{-5}$) compared to deletion of RPN4 (Figure 4.7C), thereby confirming the model prediction. Taken together, these results provide a likely mechanistic explanation where Pdr3 acts as the intermediate buffering gene in regulating Pdr5 and Pdr15 (Figure 4.7D).

4.3 Discussion

4.3.1 Genome-wide gene expression measurements to investigate the genetic interaction landscape

To investigate genetic interactions in a high-throughput manner, growth-based assays have frequently been deployed, resulting in the identification of an overwhelming number of both negative and positive genetic interactions (Bandyopadhyay et al., 2010; Costanzo et al., 2016; Davierwala et al., 2005; Fiedler et al., 2009; Jasnos and Korona, 2007; Pan et al., 2006; St Onge et al., 2007; Szappanos et al., 2011; Tong et al., 2004; Zheng et al., 2010). Based on these surveys, several theoretical mechanisms have been proposed to explain genetic interactions (Baryshnikova et al., 2013; Boucher and Jenna, 2013; Dixon et al., 2009; Lehner, 2011). More efforts, also using different types of assays, are however still needed to systematically and thoroughly investigate the underlying mechanisms. Alongside growth-based genetic interactions, genome-wide gene expression measurements have been applied to elucidate potential molecular mechanisms underlying genetic interactions (Capaldi et al., 2008; Dixit et al., 2016; Gutin et al., 2015; Pirkl et al., 2017; Sameith et al., 2015; van Wageningen et al., 2010). Although more laborious, expression-based genetic interactions potentially allow for more in-depth characterization of the genetic interaction landscape. Here, we show that buffering is the most frequently occurring pattern underlying most negative genetic interactions. These are however to a large degree related to slow growing strains, hindering the investigation of the underlying mechanisms. By applying a slow growth transformation that removes a cell cycle associated gene expression signature, many such effects can be filtered out (O'Duibhir et al., 2014). The transformation results in distinct clusters that can be more easily aligned with potential underlying mechanisms. Recent advances using Crispr-Cas9 single and double knock-down screens, followed by single cell RNA sequencing have also shown that results are greatly influenced by the cell-cycle phase in which different cells are found (Adamson et al., 2016; Dixit et al., 2016). It is therefore essential for future studies on genetic interactions to incorporate methods that decompose such large confounding effects, as they greatly influence the ability to deduce mechanism.

4.3.2 Systematic modeling to understand mechanisms of genetic interactions

To infer underlying mechanisms from the genetic interaction landscape as obtained from genome-wide gene expression measurements, systematic modeling approaches are warranted (Baryshnikova et al., 2013; Lehner, 2011). Various modeling techniques have been instrumental in understanding various aspects of experimental data (reviewed in (Karlebach and Shamir, 2008)). Different modeling methods have different applications, depending on the question asked and available data types. To infer the underlying mechanisms for many genetic interactions, an approach is needed that is able to exhaustively explore the complete genetic interaction landscape while at the same time incorporating (semi-) quantitative values. Here, using Petri net modeling, we have been able to exhaustively explore more than nine million models that included semi-quantitative effects. Inversion, a pattern strongly associated with a group of GSTF pairs was investigated in more detail, resulting in the striking conclusion that a quantitative difference in interaction strength is needed to explain inversion. The approach taken here, by combining slow growth corrected

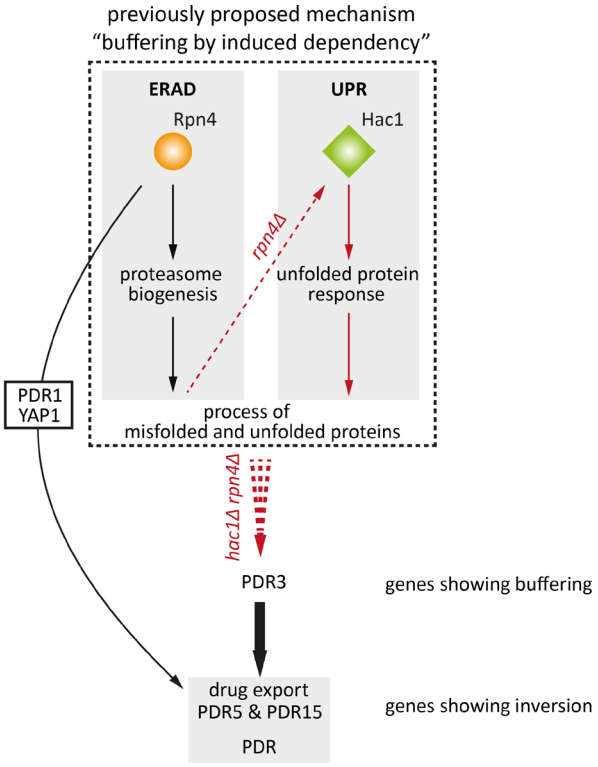


Figure 4.8. Combination of buffering by induced dependency and proposed model for inversion. Carton depiction of proposed model for genetic interaction between Rpn4 and Hac1. Red arrows indicate the consequence of disrupted genes and pathways. The dashed rectangle indicates a previously proposed model, "buffering by induced dependency", to explain genes showing buffering for Hac1-Rpn4 . A thicker arrow represents a stronger activation strength.

genome-wide gene expression measurements with the exhaustive semi-quantitative Petri-net modeling thus highlights the benefits of using such an approach to understand mechanisms of genetic interactions. Applying this approach to other types of genetic interactions or across many more genetic interaction pairs can help us in further characterizing mechanisms of genetic interactions and relating these to pathway organization and cellular states.

4.3.3 Inversion as a way to fine-tune different efficacies between two redundant processes and a third, compensatory process

Previously, a mechanism termed "buffering by induced dependency" was proposed to explain parts of the genetic interaction patterns observed between Rpn4 and Hac1 (Figure 4.8, dotted inset) (Sameith et al., 2015). This mechanism links the endoplasmic reticulum-associated degradation (ERAD) by the proteasome (Rpn4) with the unfolded protein response (UPR, Hac1), two distinct processes dealing with misfolded and unfolded proteins. By combining the "buffering by induced dependency" mechanism with the model proposed for inversion here, most genetic interaction patterns observed for Rpn4 and Hac1 can be explained (Figure 4.7A; 4.8). The combined model introduces a third, compensatory process, the pleiotropic drug response (PDR; Figure 4.8, bottom light gray inset). Even though the exact relationship between ERAD, UPR and pleiotropic drug response is unclear, the interplay between UPR and drug export has been shown in mammalian cells (Yan et al., 2015). In yeast, Pdr5 and Pdr15 have been implicated in cellular detoxification (Mamnun et al., 2004; Wolfger et al., 1997) and may also be required for cellular detoxification under normal growth conditions (Mamnun et al., 2004). Both Pdr5 and Pdr15 have been reported to be regulated through Pdr1 and Yap1 (DeRisi et al., 2000; Miyahara et al., 1996), as well as through Rpn4 (Salin et al., 2008; Spasskaya et al., 2014). This is also confirmed here by downregulation of both Pdr1 and Yap1 as well as downregulation of their target genes Pdr5 and Pdr15 in *rpn4Δ* (Figure 4.7B, C). It is therefore likely that in the wildtype situation when Rpn4 is active, both ERAD and the PDR are functioning (Figure 4.8). Deletion of RPN4 leads to deactivation of the ERAD and PDR pathways and activation of the UPR through Hac1 (Figure 4.8, *rpn4Δ* dotted red line). Deletion of both RPN4 and HAC1 results in a major growth defect and accumulation of misfolded and unfolded proteins, most likely leading to a stronger activation of the PDR through Pdr3 compared to the wildtype situation (Figure 4.7B, C; Figure 4.8, *hac1Δ rpn4Δ* dotted red line) (Katzmann et al., 1994; Wolfger et al., 1997). Taken together, this model thus provides a potential regulatory mechanism in which two redundant processes, each with slightly different efficacies, can be differentially regulated, or fine-tuned, through a third, compensatory process. The requirement to fine-tune slightly different efficacies of different cellular processes then also provides a potential explanation why inversion is observed more frequently for gene-specific transcription factors since these allow for more fine-grained control than protein kinases and phosphatases.

4.3.4 Conclusions

In conclusion, we have shown how exhaustive exploration of regulatory networks can be used to generate plausible hypothetical regulatory mechanisms underlying inversion. Almost all models showing inversion contain a quantitative difference in interaction strength and this difference is frequently accompanied with a second gene showing buffering. These hypothetical mechanisms have subsequently been tested against known and new experimental data. For GSTFs we show a validated example of Hac1-Rpn4 where differential regulation of gene expression is key to understanding the genetic interaction pattern inversion. Furthermore, the ability of GSTFs to differentially regulate expression of downstream target genes provides a likely explanation why inversion is more prevalent in GSTFs.

4.4 Materials and methods

4.4.1 Selection of GSTF and kinase/phosphatase pairs

Two selection criteria were applied to select genetically interacting GSTF and kinase/phosphatase pairs. First, one of the mutants of each individual pair should show genome-wide gene expression measurements similar to wildtype (WT). DNA microarray data from Kemmeren et al. (2014) was used to determine whether a single deletion mutant is similar to WT. A deletion mutant is considered similar to WT when fewer than eight genes are changing significantly ($p < 0.05$, FC > 1.7) in the deletion mutant gene expression profile, as previously described (van Wageningen et al., 2010). Second, selected pairs should show a significant growth-based negative genetic interaction score. Growth-based genetic interaction scores for GSTF (Zheng et al., 2010) and kinase/phosphate (Fiedler et al., 2009) pairs were converted to Z-scores. A negative Z-score significance of $p < 0.05$ after multiple testing correction was used as the significance threshold. Applying these selection criteria resulted in 11 GSTF pairs and 15 kinase/phosphatase pairs (S4.1 Table).

4.4.2 Genome-wide gene expression measurements and statistical analyses

Genome-wide gene expression measurements of single and double mutant GSTF pairs were obtained from Sameith et al. (2015). Genome-wide gene expression measurements of single and double mutant kinase/phosphatase pairs were obtained from van Wageningen et al. (2010). Genome-wide gene expression measurements of *pdr3Δ* and *rpn4Δ* were obtained from Kemmeren et al. (2014). Statistical analysis of these gene expression profiles was performed as previously described (Kemmeren et al., 2014). In summary, mutants were grown in Synthetic Complete (SC) medium with 2% glucose and harvested during exponential growth. WT cultures were grown alongside mutants in parallel to monitor for day to day effects. For each mutant statistical analysis using limma was performed versus a collection of WTs (Kemmeren et al., 2014; van Wageningen et al., 2010). Reported FC for each transcript is the average of four replicate expression profiles over a WT pools consisting of 200 WT strains.

4.4.3 Growth-based genetic interaction scores

Growth measurements for single and double mutant GSTF and kinase/phosphatase pairs were obtained from Sameith et al. (2015) and van Wageningen et al. (2010) respectively. Growth-based genetic interaction scores were calculated for both GSTF and kinase/phosphatase pairs as performed before (Sameith et al., 2015). In summary, the fitness W of single and double mutants was determined as the ratio between the WT growth rate and the mutant growth rate. The growth-based genetic interaction score $\varepsilon_{growth,XY}$ was calculated as the deviation of the observed fitness in a double mutant from the expected fitness based on the respective single mutants ($\varepsilon_{growth,XY} = W_{xAyA} - W_{xA} \cdot W_{yA}$). p values were assigned to genetic interaction scores based on the mean and standard deviation of a generated background distribution (Sameith et al., 2015). p values were corrected for multiple testing using Benjamini-Hochberg. Adjusted p values lower than 0.05 were considered significant. Fitness values of all single and double mutants, as well as calculated genetic interaction scores can be found in S4.1 Table.

4.4.4 Expression-based genetic interaction scores

Expression-based genetic interaction scores were calculated for both GSTF and kinase/phosphatase pairs as described before (Sameith et al., 2015). In summary, the effect of a genetic interaction between two genes X and Y on gene i is calculated as the deviation between the observed expression change in the double mutant and the expected expression change based on the corresponding single mutants ($\varepsilon_{txpn\_i,XY} = |M_{i,xAyA} - (M_{i,xA} + M_{i,yA})|$). The overall genetic interaction score between X and Y is calculated as the sum all genes i for which $\varepsilon_{txpn\_i,XY} > \log_2(1.5)$. All genetic interaction scores consisting of at least 10 genes were kept for further downstream analyses. Genes with similar gene expression changes were divided into the 6 different patterns (buffering, quantitative buffering, suppression, quantitative suppression, masking, inversion), as previously described (Sameith et al., 2015) (Figure 4.1A).

4.4.5 Clustering of expression-based genetic interaction scores

Genetic interaction profiles for both classes of proteins were grouped together based on the number of occurrences of the six different patterns using hierarchical clustering. Average linkage was applied for the clustering. Identity of genes in each genetic interaction profile was disregarded.

4.4.6 Slow growth transformation

Slow growth signature transformation of the gene expression profiles was performed as previously described (O'Duibhir et al., 2014). In short, for each mutant, the correlation of its expression profile with the first principal component of 1,484 deletion strains (Kemmeren et al., 2014) was removed, thus minimizing correlation with the relative growth rate. The transformation reduces correlation with the relative growth rate from 0.29 to 0.10 on average (O'Duibhir et al., 2014).

4.4.7 Model generation

Exhaustive modeling of possible network topologies underlying the genetic interaction patterns was carried out by creating Petri net models consisting of four nodes, representing two regulator genes (R1 and R2) and two downstream genes (G1 and G2). With four nodes and directed edges, there are $4^2 = 16$ possible edges, and $2^{16} = 65536$ possible edge weight matrices, which is a tractable number. However, each interaction can in addition be positive or negative, and weak or strong (and absent), leading to $5^{16} = 1.5 \times 10^{11}$ possible interaction graphs (edge weight matrices), which becomes intractable. Many of these models, however, will be irrelevant for the understanding the biological behavior of genetic interaction patterns of two genes. To exclude these types of models, the following conditions were applied: 1) No self-edges are allowed. 2) The number of incoming edges on any node must be limited to two. 3) At least two incoming edges from at least one of the regulators (upstream nodes) to the genes (downstream nodes). Applying these conditions reduces the number of relevant edge weight matrices to 9,287,616. Furthermore, most generated matrices have mirror counterparts, therefore only one of the matrices was included in downstream analyses. Applying this filtering step results in 2,323,936 matrices. Figure 4.3 gives an overview of the various filtering steps, and shows which representation of the models was relevant in different stages of the filtering. Edge weight matrices were generated in R, version 3.2.2 (the function expand.grid was used to generate all combinations of edges per row in a given matrix).

4.4.8 Petri net simulations

Regulatory effects of two potentially interacting genes (R1 and R2) on two downstream genes (G1 and G1) were simulated using a Petri net approach (Bonzanni et al., 2009; Bonzanni et al., 2009; Jacobsen et al., 2016; Krepska et al., 2008; Petri, 1962) to recapitulate genetic interaction patterns observed in the gene expression data.

In the Petri net notation, nodes in a given model are represented by places (denoted as circles). Interactions between nodes always go via a transition (denoted as squares), connected via directed arcs (drawn as arrows). An incoming arc to a transition can be either activating or inhibiting. The weight on arcs going to a transition is always fixed to 1. The weight on arcs going from a transition to a place depends on the edge weight between two nodes, 1 for weak and 5 for strong (Figure 4.3). For nodes with two incoming edges, one has to decide how these two inputs should be combined: does the transition require both inputs to be activated (AND logic), or can one or the other activate it (OR logic). To incorporate this, for each pair of incoming edges with the same weight, two Petri net models were generated: one using the AND logic, and one using the OR logic (Figure 4.3, bottom right panel). For two incoming edges with different weights only the Petri net model using the OR logic was generated. For cases with two incoming edges to a node with two different directions, activation and inhibition, inhibition dominates.

To simulate the regulatory effects of two upstream genes (R1 and R2), 200 tokens were provided to represent the mRNA resources for each regulator, except when one of the regulators has an incoming edge from the other regulator as shown in (S4.5A Figure). Each step in the simulation process comprises of firing all enabled transitions (maximal parallel execution) (Bonzanni et al., 2009; Burkhard, 1980). A transition is enabled to fire when resources (tokens) in the input place(s) match or exceed the weight(s) on the respective incoming arc(s) to the transition (S4.5B Figure). In total 50 consecutive transition firing steps were performed.

To incorporate deletion mutants in the simulation process, tokens were removed from corresponding regulators. To prevent accumulation of tokens in deleted regulators, each outgoing arc from a transition to the corresponding deleted places were also removed in simulated deletion strains. The number of tokens in G1 and G2 after 50 steps of firing transitions in single and double mutants were compared with that in the WT situation where both R1 and R2 are active. To avoid division by zero one token was added to the total number of tokens in G1 and G1. These fold changes were then $\log_2$ transformed (M values).

Simulation-based genetic interaction scores for G1 and G2 were calculated based on the deviation between observed M values in the double mutant and the expected M value based on the single mutants, as follows: $\varepsilon_{sim,R1R2i} = |M_{R1\Delta R2\Delta i} - (M_{R1\Delta i} + M_{R2\Delta i})|$, where *i* can be either G1 or G2. Each node with $\varepsilon_{sim,R1R2i} > \log2(1.7)$ was further divided into genetic interaction patterns, as defined before based on gene expression data (Sameith et al., 2015). Simulated expression levels for single and double mutants are considered to be increased relative to WT when $M > \log_2(1.7)$ and decreased when $M < -\log_2(1.7)$.

4.4.9 Functional enrichment test

Functional enrichment analyses were performed using a hypergeometric testing procedure on Gene Ontology (GO) biological process (BP) annotations (Butow and Avadhani, 2004) obtained from the *Saccharomyces cerevisiae* Database (Giannattasio et al., 2005). The background population of genes was set to 6,359 and *p* values were corrected for multiple testing using Bonferroni.

4.4.10 Visualization of models

Models were visualized in R, version 3.2.2, using diagram package (version 1.6.3). Weak and strong activation/inhibition edges are represented as thin and thick lines, respectively.

References

Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. 2016, 167, 1867-82.e21.

Amini S, Holstege FCP, Kemmeren P. Growth condition dependency is the major cause of non-responsiveness upon genetic perturbation. *PloS One*. 2017, 12, e0173432.

Babu M, Arnold R, Bundalovic-Torma C, Gagarinova A, Wong KS, Kumar A, et al. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in Escherichia coli. *PLoS Genet*. 2014, 10, e1004120.

Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet*. 2002, 3, 779-89.

Bakal C, Linding R, Llense F, Heffern E, Martin-Blanco E, Pawson T, et al. Phosphorylation Networks Regulating JNK Activity in Diverse Genetic Backgrounds. *Science*. 2008, 322, 453-6.

Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, et al. Rewiring of genetic networks in response to DNA damage. *Science*. 2010, 330, 1385-9.

Baryshnikova A, Costanzo M, Myers CL, Andrews B, Boone C. Genetic interaction networks: toward an understanding of heritability. *Annu Rev Genomics Hum Genet*. 2013, 14, 111-33.

Billmann M, Horn T, Fischer B, Sandmann T, Huber W, Boutros M. A genetic interaction map of cell cycle regulators. *Mol Biol Cell*. 2016, 27, 1397-407.

Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D. Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. *Mol Cell Biol*. 1997, 17, 3640-8.

Bonzanni N, Feenstra KA, Fokkink W, Heringa J. Petri Nets Are a Biologist's Best Friend. Formal Methods in Macro-Biology. Springer, Cham; 2014. pp. 102-16.

Bonzanni N, Feenstra KA, Fokkink W, Krepska E. What Can Formal Methods Bring to Systems Biology? FM 2009: Formal Methods. Springer, Berlin, Heidelberg; 2009. pp. 16-22.

Bonzanni N, Garg A, Feenstra KA, Schütte J, Kinston S, Miranda-Saavedra D, et al. Hardwired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013, 29, i80-8.

Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. *Bioinformatics*. 2009, 25, 2049-56.

Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. *Nat Rev Genet*. 2007, 8, 437-49.

Boucher B, Jenna S. Genetic interaction networks: better understand to better predict. *Front Genet*. 2013, 4, 290.

Bouquin N, Barral Y, Courbeyrette R, Blondel M, Snyder M, Mann C. Regulation of cytokinesis by the Elm1 protein kinase in Saccharomyces cerevisiae. *J Cell Sci*. 2000, 113, 1435-45.

Burkhard H-D. On Priorities of Parallelism. Logic of Programs and Their Applications, Proceedings. London, UK, UK: Springer-Verlag; 1983. pp. 86-97.

Butow RA, Avadhani NG. Mitochondrial Signaling: The Retrograde Response. *Mol Cell*. 2004, 14, 1-15.

Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nat Genet*. 2008, 40, 1300-6.

Chaouiya C, Remy E, Thieffry D. Qualitative Petri Net Modelling of Genetic Networks. Transactions on Computational Systems Biology VI. Springer, Berlin, Heidelberg; 2006. pp. 95-112.

Coffman JA, Rai R, Cunningham T, Svetlov V, Cooper TG. Gat1p, a GATA family protein whose production is sensitive to nitrogen catabolite repression, participates in transcriptional activation of nitrogen catabolic genes in Saccharomyces cerevisiae. *Mol Cell Biol*. 1996, 16, 847-58.

Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*. 2007, 446, 806-10.

Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet*. 2013, 132, 1077-130.

Cope MJ, Yang S, Shang C, Drubin DG. Novel protein kinases Ark1p and Prk1p associate with and regulate the cortical actin cytoskeleton in budding yeast. *J Cell Biol*. 1999, 144, 1203-18.

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science*. 2016, 353, aaf1420.

Costigan C, Kolodrubetz D, Snyder M. NHP6A and NHP6B, which encode HMG1-like proteins, are candidates for downstream components of the yeast SLT2 mitogen activated protein kinase pathway. *Mol Cell Biol*. 1994, 14, 2391-403.

Cunningham TS, Cooper TG. Expression of the DAL80 gene, whose product is homologous to the GATA factors and is a negative regulator of multiple nitrogen catabolic genes in Saccharomyces cerevisiae, is sensitive to nitrogen catabolite repression. *Mol Cell Biol*. 1991, 11, 6205-15.

Dagsgaard C, Taylor LE, O'Brien KM, Poyton RO. Effects of anoxia and the mitochondrion on expression of aerobic nuclear COX genes in yeast: evidence for a signaling pathway from the mitochondrial genome to the nucleus. *J Biol Chem*. 2001, 276,7593-601.

Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, et al. The synthetic genetic interaction spectrum of essential genes. *Nat Genet*. 2005, 37, 1147-52.

DeRisi J, van den Hazel B, Marc P, Balzi E, Brown P, Jacq C, et al. Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. *FEBS Lett*. 2000, 470, 156-60.

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 2016, 167, 1853-66.e17.

Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. Systematic mapping of genetic interaction networks. *Annu Rev Genet*. 2009, 43, 601-25.

Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional Organization of the S. cerevisiae Phosphorylation Network. *Cell*. 2009, 136, 952-63.

Giannattasio S, Liu Z, Thornton J, Butow RA. Retrograde Response to Mitochondrial Dysfunction Is Separable from TOR1/2 Regulation of Retrograde Gene Expression. *J Biol Chem*. 2005, 280, 42528-35.

Golin J, Ambudkar SV, May L. The yeast Pdr5p multidrug transporter: how does it recognize so many substrates? *Biochem Biophys Res Commun*. 2007, 356, 1-5.

Gutin J, Sadeh A, Rahat A, Aharoni A, Friedman N. Condition-specific genetic interaction maps reveal crosstalk between the cAMP/PKA and the HOG MAPK pathways in the activation of the general stress response. *Mol Syst Biol*. 2015, 11, 829.

Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol*. 2017, 35, 463-74.

Hartman JL, Garvik B, Hartwell L. Principles for the buffering of genetic variation. *Science*. 2001, 291, 1001-4.

Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat Methods*. 2011, 8, 341-6.

Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Mol Syst Biol*. 2007, 3, 86.

Jacobsen A, Heijmans N, Verkaar F, Smit MJ, Heringa J, Amerongen R van, et al. Construction and Experimental Validation of a Petri Net Model of Wnt/β-Catenin Signaling. *PLoS One*. 2016, 11, e0155743. **Chapter 2 of this thesis.**

Jasnos L, Korona R. Epistatic buffering of fitness loss in yeast double deletion strains. *Nat Genet*. 2007, 39, 550-4.

Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008, 9, 770-80.

Katzmann DJ, Burnett PE, Golin J, Mahé Y, Moye-Rowley WS. Transcriptional control of the yeast PDR5 gene by the PDR3 gene product. *Mol Cell Biol*. 1994, 14, 4653-61.

Kauffman S, Peterson C, Samuelsson B, Troein C. Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci U S A*. 2003, 100, 14796-9.

Keane OM, Toft C, Carretero-Paulet L, Jones GW, Fares MA. Preservation of genetic and regulatory robustness in ancient gene duplicates of Saccharomyces cerevisiae. *Genome Res*. 2014, 24, 1830-41.

Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell*. 2014, 157, 740-52.

Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, et al. Promoters maintain their relative activity levels under different growth conditions. *Mol Syst Biol*. 2013, 9, 701.

Krepska E, Bonzanni N, Feenstra A, Fokkink W, Kielmann T, Bal H, et al. Design Issues for Qualitative Modelling of Biological Cells with Petri Nets. Proceedings of the Formal Methods in Systems Biology 2008, Vol 5054 of LNBI, Springer, Cambridge, UK. 2008, 48-62.

Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet*. 2006, 38, 896-903.

Lehner B. Molecular mechanisms of epistasis within and between genes. *Trends Genet TIG*. 2011, 27, 323-31.

Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A*. 2004, 101, 4781-6.

Mamnun YM, Schüller C, Kuchler K. Expression regulation of the yeast PDR5 ATP-binding cassette (ABC) transporter suggests a role in cellular detoxification during the exponential growth phase. *FEBS Lett*. 2004, 559, 111-7.

Mani R, St Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. *Proc Natl Acad Sci U S A*. 2008, 105, 3461-6.

Mannhaupt G, Schnall R, Karpov V, Vetter I, Feldmann H. Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Lett*. 1999, 450, 27-34.

McMillan JN, Longtine MS, Sia RA, Theesfeld CL, Bardes ES, Pringle JR, et al. The morphogenesis checkpoint in Saccharomyces cerevisiae: cell cycle control of Swe1p degradation by Hsl1p and Hsl7p. *Mol Cell Biol*. 1999, 19, 6929-39.

Minehart PL, Magasanik B. Sequence and expression of GLN3, a positive nitrogen regulatory gene of Saccharomyces cerevisiae encoding a protein with a putative zinc finger DNA-binding domain. *Mol Cell Biol*. 1991, 11, 6216-28.

Miyahara K, Hirata D, Miyakawa T. yAP-1- and yAP-2-mediated, heat shock-induced transcriptional activation of the multidrug resistance ABC transporter genes in Saccharomyces cerevisiae. *Curr Genet*. 1996, 29, 103-5.

Moore JH, Williams SM. Epistasis and its implications for personal genetics. *Am J Hum Genet*. 2009, 85, 309-20.

Morgan DO. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol*. 1997, 13, 261-91.

Mori K, Kawahara T, Yoshida H, Yanagi H, Yura T. Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. *Genes Cells Devoted Mol Cell Mech*. 1996, 1, 803-17.

O'Duibhir E, Lijnzaad P, Benschop JJ, Lenstra TL, van Leenen D, Groot Koerkamp MJA, et al. Cell cycle population effects in perturbation studies. *Mol Syst Biol*. 2014, 10, 732.

Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. A DNA integrity network in the yeast Saccharomyces cerevisiae. *Cell*. 2006, 124, 1069-81.

Petri CA. Kommunikation mit Automaten. PhD Thesis, Technische Universität Darmstadt, Bonn, Germany. 1962.

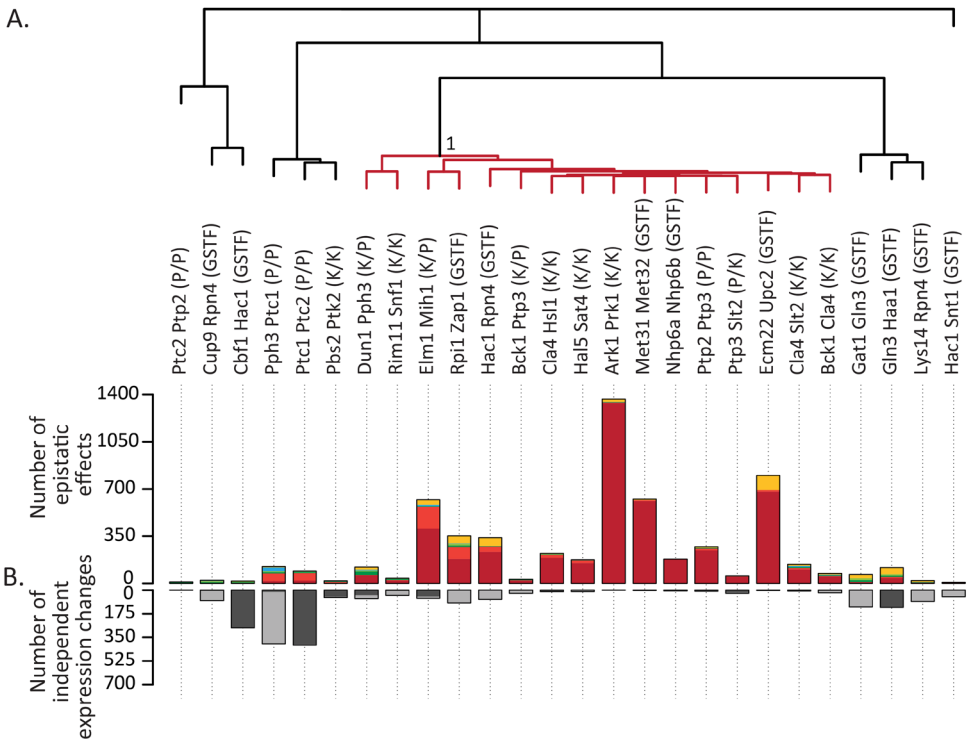
Phillips PC. Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008, 9, 855-67.

Pirkl M, Diekmann M, van der Wees M, Beerenwinkel N, Fröhlich H, Markowetz F. Inferring modulators of genetic interactions with epistatic nested effects models. *PLoS Comput Biol*. 2017, 13, e1005496.

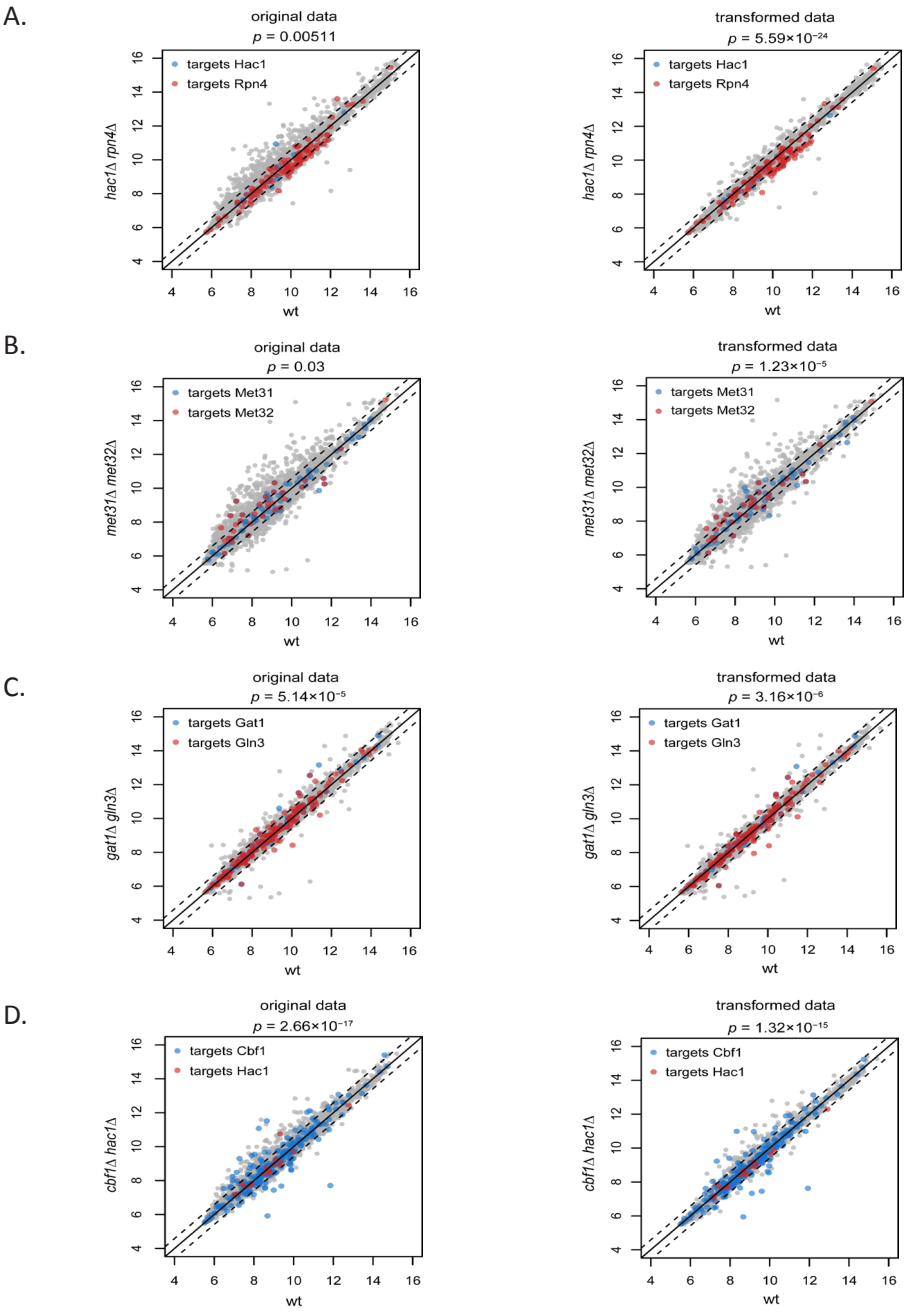
Plata G, Vitkup D. Genetic robustness and functional evolution of gene duplicates. *Nucleic Acids Res*. 2014, 42, 2405-14.
Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, Bro C, et al. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. *Genome Biol*. 2006, 7, R107.
Roguev A, Talbot D, Negri GL, Shales M, Cagney G, Bandyopadhyay S, et al. Quantitative genetic-interaction mapping in mammalian cells. *Nat Methods*. 2013, 10, 432-7.
Russell P, Moreno S, Reed SI. Conservation of mitotic controls in fission and budding yeasts. *Cell*. 1989, 57, 295-303.
Russell P. Checkpoints on the road to mitosis. *Trends Biochem Sci*. 1998, 23, 399-402.
Salin H, Fardeau V, Piccini E, Lelandais G, Tanty V, Lemoine S, et al. Structure and properties of transcriptional networks driving selenite stress response in yeasts. *BMC Genomics*. 2008, 9, 333.
Sameith K, Amini S, Groot Koerkamp MJA, van Leenen D, Brok M, Brabers N, et al. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biol*. 2015, 13, 112.
Saxena D, Kannan KB, Brandriss MC. Rapamycin Treatment Results in GATA Factor-Independent Hyperphosphorylation of the Proline Utilization Pathway Activator in Saccharomyces cerevisiae. *Eukaryot Cell*. 2003, 2, 552-9.
Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, et al. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat Methods*. 2017, 14, 573-6.
Spasskaya DS, Karpov DS, Mironov AS, Karpov VL. Transcription factor Rpn4 promotes a complex antistress response in Saccharomyces cerevisiae cells exposed to methyl methanesulfonate. *Mol Biol*. 2014, 48, 141-9.
St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat Genet*. 2007, 39, 199-206.
Stanbrough M, Rowen DW, Magasanik B. Role of the GATA factors Gln3p and Nil1p of Saccharomyces cerevisiae in the expression of nitrogen-regulated genes. *Proc Natl Acad Sci U S A*. 1995, 92, 9450-4.
Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet*. 2011, 43, 656-62.
Thomas CL, Blacketer MJ, Edgington NP, Myers AM. Assembly interdependence among the S. cerevisiae bud neck ring proteins Elm1p, Hsl1p and Cdc12p. *Yeast Chichester Engl*. 2003, 20, 813-26.
Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. *Science*. 2004, 303, 808-13.
van Wageningen S, Kemmeren P, Lijnzaad P, 744 Margaritis T, Benschop JJ, de Castro IJ, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell*. 2010, 143, 991-1004.
Vik A, Rine J. Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae. *Mol Cell Biol*. 2001, 21, 6395-405.
Vizeacoumar FJ, Arnold R, Vizeacoumar FS, Chandrashekhar M, Buzina A, Young JTF, et al. A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol*. 2013, 9, 696.
Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet. 2014, 15, 722-33.
Wolfger H, Mahé Y, Parle-McDermott A, Delahodde A, Kuchler K. The yeast ATP binding cassette (ABC) protein genes PDR10 and PDR15 are novel targets for the Pdr1 and Pdr3 transcriptional regulators. *FEBS Lett*. 1997, 418, 269-74.
Wong ASL, Choi GCG, Lu TK. Deciphering Combinatorial Genetics. *Annu Rev Genet*. 2016, 50, 515-38.
Yan MM, Ni JD, Song D, Ding M, Huang J. Interplay between unfolded protein response and autophagy promotes tumor drug resistance. *Oncol Lett*. 2015, 10, 1959-69.
Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, et al. Epistatic relationships reveal the functional organization of yeast transcription factors. *Mol Syst Biol*. 2010, 6, 420.

S4.1 Table. Single and double mutant GSTF and kinase/phosphatase pairs.

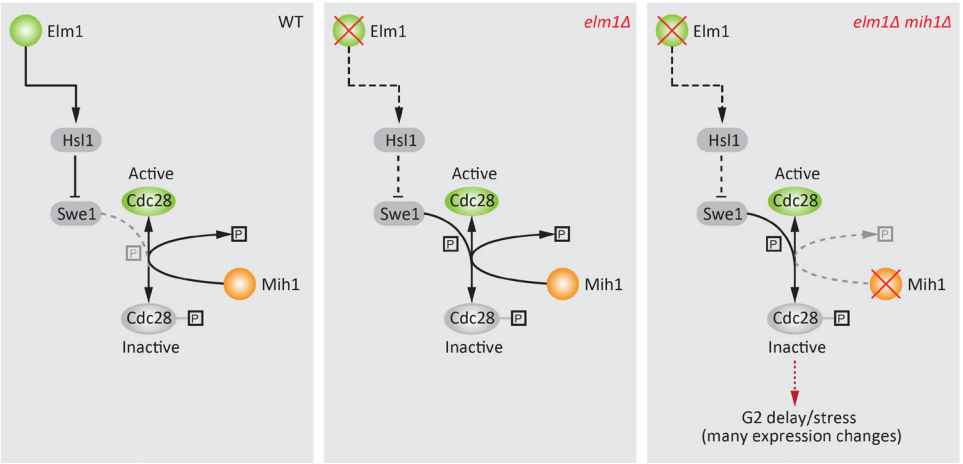
Gene pair	GSTF or K/P	RGR (observed)	RGR (expected)	RGR (single mutants)	genetic interaction score	adjusted *p*
Met31 / Met32	GSTF	0.296	1.059	1.031 / 1.028	-0.763	2.05×10^{-21}
Ecm22 / Upc2	GSTF	0.506	1.040	1.005 / 1.034	-0.534	8.17×10^{-11}
Hac1 / Rpn4	GSTF	0.453	0.958	1.056 / 0.907	-0.504	8.76×10^{-10}
Nhp6a / Nhp6b	GSTF	0.455	0.891	0.923 / 0.965	-0.436	1.12×10^{-7}
Gln3 / Haa1	GSTF	0.640	0.861	0.803 / 1.073	-0.221	0.01340
Rpi1 / Zap1	GSTF	0.621	0.824	0.965 / 0.854	-0.203	0.02434
Cbf1 / Hac1	GSTF	0.603	0.720	0.682 / 1.056	-0.117	0.19109
Cup9 / Rpn4	GSTF	0.924	1.022	1.127 / 0.907	-0.098	0.25171
Hac1 / Snt1	GSTF	0.902	0.951	1.056 / 0.901	-0.049	0.38827
Gat1 / Gln3	GSTF	0.780	0.818	1.019 / 0.803	-0.039	0.41165
Lys14 / Rpn4	GSTF	1.036	0.799	0.881 / 0.907	0.2376	0.00397
Pbs2 / Ptk2	K / K	0.840	0.889	0.949 / 0.937	-0.049	0.38827
Ptc2 / Ptp2	P / P	1.003	0.989	1.006 / 0.983	0.014	0.43533
Pph3 / Ptc1	P / P	0.412	0.564	0.967 / 0.583	-0.152	0.09558
Bck1 / Cla4	K / K	0.431	0.874	0.973 / 0.898	-0.443	7.61×10^{-8}
Hal5 / Sat4	K / K	0.718	0.921	0.980 / 0.940	-0.203	0.02434
Elm1 / Mih1	K / P	0.316	0.816	0.849 / 0.961	-0.500	9.78×10^{-10}
Ark1 / Prk1	K / K	0.412	0.976	0.989 / 0.986	-0.563	7.64×10^{-12}
Bck1 / Ptp3	K / P	0.810	0.954	0.973 / 0.980	-0.143	0.11371
Ptc1 / Ptc2	P / P	0.538	0.587	0.583 / 1.006	-0.048	0.38827
Rim11 / Snf1	K / K	0.874	0.942	0.977 / 0.965	-0.068	0.33889
Dun1 / Pph3	K / P	0.785	0.731	0.755 / 0.967	0.054	0.33889
Cla4 / Slt2	K / K	0.496	0.865	0.898 / 0.964	-0.369	0.0000107
Ptp3 / Slt2	P / K	0.492	0.944	0.980 / 0.964	-0.453	4.11×10^{-8}
Ptp2 / Ptp3	P / P	0.712	0.963	0.983 / 0.980	-0.251	0.0039742
Cla4 / Hsl1	K / K	0.612	0.882	0.898 / 0.983	-0.270	0.0023007



S4.1 Figure. Buffering dominates genetic interaction profiles. A) Hierarchical clustering of all pairs according to their genetic interaction effects. Average linkage clustering was applied to group pairs with similar genetic interaction patterns. The number of occurrences for each genetic interaction pattern was used and the identity of individual genes was disregarded. Similarity between pairs was calculated using the cosine correlation. Most pairs are grouped together in a single branch (indicated in red), which is dominated by buffering. B) The number of genetic interaction effects underlying the clustering are shown as bar plots below the dendrogram (top; colors as in Fig 1A). B) Number of genes showing no genetic interaction pattern but significantly changing in one of the mutants compared to WT (bottom; $p \leq 0.01$, FC > 1.5). Dark gray for the first named gene, light gray for the second named gene.



S4.2 Figure. Slow growth correction improves identification of GSTF targets. Scatter plots showing gene expression levels in the GSTF double mutant pairs *hac1Δ rpn4Δ* A), *met31Δ met32Δ* B), *gat1Δ gln3Δ* C) and *cbf1Δ hac1Δ* D) versus WT before (left) or after (right) slow growth correction. Individual transcripts are represented as dots. The dashed line indicates a FC of 1.7. Dots depicted in blue and red correspond to targets of the first and second gene in a named GSTF pair. *p* values are calculated using a hypergeometric testing procedure to test the enrichment of GSTF targets among genes that change more than 1.7 fold before (left) or after (right) slow growth correction.



S4.3 Figure. The genetic interaction between Elm1 and Mih1 can be explained through pathway redundancy. Cartoon depicting the proposed genetic interaction between Elm1 and Mih1. (left panel) WT situation where the activity of Cdc28 is not disrupted by Swe1 phosphorylation. (Middle panel) Deletion of Elm1 leads to derepression of Swe1 activity. The increase of Swe1 activity can be compensated by Mih1. (Right panel) Deletion of both Elm1 and Mih1 will cause an increase of phosphorylated Cdc28 (inactive form), which in turn can lead to G2 delay/stress and therefore many gene expression changes.

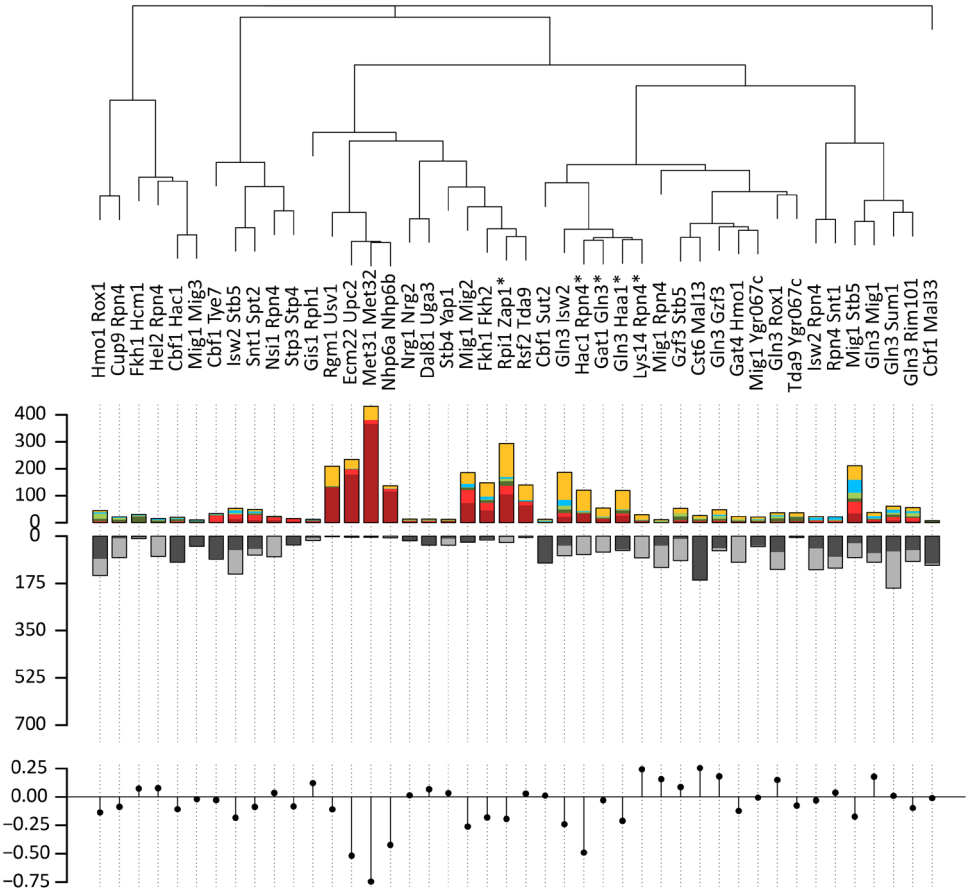
S4.2 Table. Number of genes for each genetic interaction pattern for both GSTF as well as kinases/ phosphatase pairs.

gene pair	Cbf1/ Hac1	Gat1/ Gln3	Gln3/ Haa1	Lys14/ Rpn4	Met31/ Met32	Nhp6a/ Nhp6b	Rpi1/ Zap1	Cup9/ Rpn4	Hac1/ Rpn4	Hac1/ Snt1	Ecm22/ Upc2
class	GSTF	GSTF	GSTF	GSTF	GSTF	GSTF	GSTF	GSTF	GSTF	GSTF	GSTF
buffering	3	13	27	5	366	116	105	3	33	7	180
inversion	3	30	69	19	51	11	125	1	76	2	35
q. buffering	2	1	7	1	14	9	32	0	1	2	18
suppression	8	4	12	3	0	0	17	8	4	3	1
q. suppression	1	4	3	1	0	0	9	7	3	1	0
masking	3	2	1	1	0	1	6	3	3	1	1

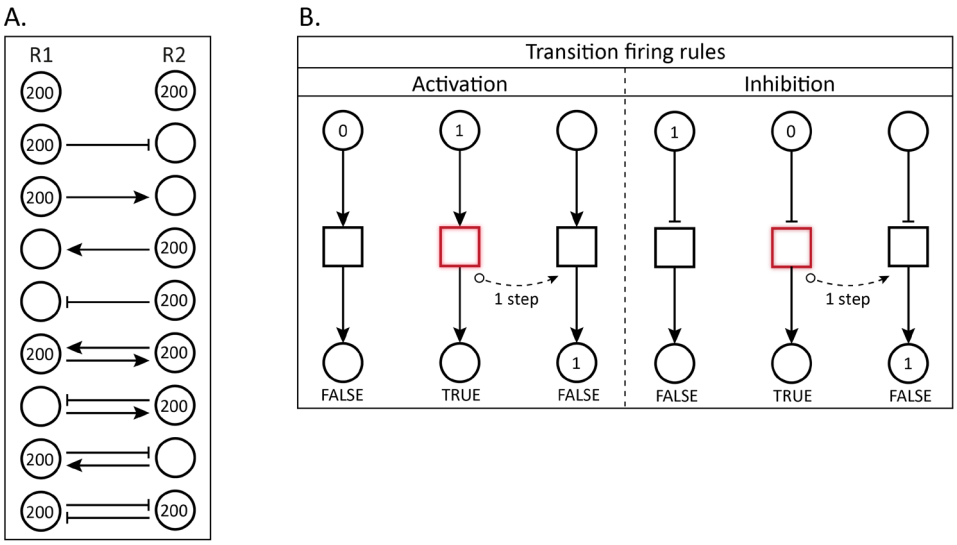
gene pair	overall # of genes	overall % of genes
class		
buffering	858	58.1
inversion	422	28.6
q. buffering	87	5.9
suppression	60	4.1
q. suppression	29	2
masking	22	1.5

gene pair	Ark1/ Prk1	Bck1/ Cla4	Bck1/ Ptp3	Dun1/ Pph3	Hal5/ Sat4	Cla4/ Hsl1	Elm1/ Mih1	Pbs2/ Ptk2	Pph3/ Ptc1	Ptc1/ Ptc2	Ptc2/ Ptp2
class	K/K	K/K	K/P	K/P	K/K	K/K	K/P	K/K	P/P	P/P	P/P
buffering	400	15	10	57	72	42	332	1	13	28	2
inversion	38	28	5	37	6	11	57	4	10	12	1
q. buffering	16	2	3	7	58	11	88	8	7	9	0
suppression	4	8	3	20	2	9	8	2	9	10	13
q. suppression	1	10	1	5	1	0	6	1	9	5	0
masking	2	0	3	6	1	2	1	1	12	2	1

gene pair	Ptp2/ Ptp3	Cla4/ Slt2	Ptp3/ Slt2	Rim11/ Snf1	overall # of genes	overall % of genes
class	P/P	K/K	P/K	K/K		
buffering	55	67	21	19	1134	61.1
inversion	24	28	10	6	277	14.9
q. buffering	15	10	2	5	241	13
suppression	9	8	4	8	117	6.3
q. suppression	0	13	2	2	56	3
masking	0	0	0	1	32	1.7



S4.4 Figure. Hierarchical clustering of positive and negative genetic interaction GSTF pairs. Hierarchical clustering of 44 GSTF pairs according to their genetic interaction effects after slow growth correction. These pairs include both negative and positive genetic interactions. Layout and analysis similar to Fig 2.



S4.5 Figure. Provided tokens to places in WT condition and transition firing rules. A) Provided tokens to regulators depending on edges between them. B) Transition firing rules for activation and inhibition edges depending on the presence of tokens in upstream places.

CHAPTER 5

A framework for exhaustive modeling of genetic interaction patterns using Petri nets

Submitted

Annika Jacobsen*, Olga Ivanova*, Saman Amini*, Patrick Kemmeren, Jaap Heringa, K. Anton Feenstra
* shared first authors

Abstract

Genetic interaction patterns are characterized by the phenotypes of interacting single and double mutated gene pairs. Uncovering the regulatory mechanisms of genetic interactions would provide a better understanding of their role in biological processes, diseases, and drug response. For a defined set of factors computational analyses can provide insights into the underpinning mechanisms of genetic interactions.

In this study, we present a framework for exhaustive modeling of genetic interaction patterns using Petri nets. Four-node models were defined with restrictions and generated on three levels to enable an exhaustive approach. Simulations suggest ~5 million models of genetic interactions for further analysis. We propose putative mechanisms for the genetic interaction patterns, inversion and suppression, by using generalized topologies and frequent edges and edge weights. Our results demonstrate that exhaustive Petri net modeling can be applied to reason about mechanisms of genetic interactions when only the phenotypes of the interacting gene pairs are known. The framework can be applied to other genetic interaction or genetic regulatory datasets.

The framework is available at http://www.ibi.vu.nl/programs/ExhMod.

5.1 Introduction

The phenotype caused by a gene mutation can be co-regulated by another mutation. This interaction between mutated gene pairs is referred to as a genetic interaction or epistasis, and occurs when the phenotype of mutating both genes differs from the expected, combined phenotype of mutating each gene individually (Phillips, 2008). Genetic interactions can be detected using readouts such as survival (Tong et al., 2004), growth (Costanzo et al., 2011), or gene expression (Roberts et al., 2000) and can broadly be divided in positive and negative genetic interactions, where the phenotype of the double mutant is either better or worse than expected compared to the phenotype of the single mutants. In addition, gene expression based genetic interactions provide further subdivision into more specific genetic interaction patterns including buffering, quantitative buffering, suppression, quantitative suppression, masking and inversion, and could also contribute to understanding mechanism (Sameith et al., 2015).

Genetic interactions are highly prevalent in genetic networks (Lehner et al., 2006; Babu et al., 2014; Costanzo et al., 2016) and play a crucial role in complex human diseases (Mackay and Moore, 2014). Nevertheless, regulatory mechanisms underlying most genetic interactions are poorly characterized (Lagator et al., 2017; Large et al., 2017; van Leeuwen et al., 2017). These mechanisms can, on a molecular level, be described by factors involved in the interconnection between two interacting genes and a target gene; thus, there are multiple mechanisms possible for each genetic interaction. In addition, more genes could affect the genetic interaction shown in the target gene. While the combinations of molecular mechanisms between three components can be reasoned intuitively, adding even one additional gene makes this impossible. The multitude of possibilities to obtain the same behavior limits the reasoning and generalization over these mechanisms. Computational approaches can help to overcome these restrictions and build a comprehensive view of mechanisms underlying genetic interactions.

Genetic interactions have previously been studied using Boolean models (van Wageningen et al., 2010), Bayesian networks (Pe'er, 2005), Nested effect models with logical functions (Pirkl et al., 2017), and Petri nets (Mayo and Beretta, 2011). Petri nets are a well suited modeling formalism to simulate the dynamic behavior of biological networks, as it is easy to implement, uses coarse-grained data, and can capture both qualitative and quantitative traits (Chaouiya, 2007; Bonzanni et al., 2009; Bonzanni et al., 2013; Jacobsen et al., 2016).

In this study, we describe a framework for exhaustive modeling of genetic interaction patterns using Petri nets. We first describe how the models are defined and created, then how specific conditions are implemented and simulated using Petri nets, and finally how we analyze the simulation results and map them to genetic interaction patterns. We have applied the same framework in a separate study (Amini et al., 2018) where we investigated inversion in gene specific transcription factors (GSTFs) and signaling genes (kinases and phosphatases) in *Saccharomyces cerevisiae*.

Here, we applied our framework to investigate factors defining genetic interaction patterns. We then further explored inversion and suppression, which are not well understood, to provide insights into their putative regulatory mechanisms, while not being restricted by strong assumptions on the nature of those mechanisms.

5.2 Materials and methods

5.2.1 Overview of the framework

We developed a framework for exhaustive modeling of genetic interaction patterns using Petri nets, following the conceptual framework of Haydarlou et al. (2016), consisting of, 1) model definition and generation, 2) model initiation and simulation, 3) assigning patterns to models, and 4) downstream analyses (Figure 5.1). These parts are further divided into 10 subparts. 1A) Models for genetic interactions are formally defined including their restrictions. 1B) Boolean adjacency matrices (BAMs) are generated. 1C) Weighted edge matrices (WEMs) are generated. 1D) Symmetrical WEMs are removed. 1E) Petri net models are generated. 2A) Petri nets are initiated for four conditions: wild type (WT), two single and one double deletion mutant. 2B) Petri nets are simulated. 3A) Genetic interaction patterns are assigned to the models based on the simulation results in 2B. 2B and 3A are iterated five times. 3B) 'Unstable' models (i.e., models not assigned to the same pattern for each simulation) and models with rare patterns (as defined by Sameith et al. (2015)) are removed. 4) Downstream analyses on models with assigned genetic interaction patterns are conducted to suggest generalized mechanisms.

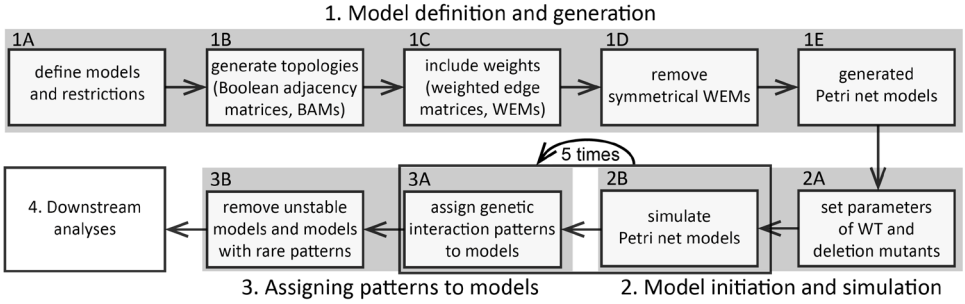


Figure 5.1. Overview of the framework for exhaustive modeling of genetic interaction patterns using Petri nets. The framework consists of four main parts: 1) model definition and generation, 2) model initiation and simulation, 3) assigning patterns to models, and 4) downstream analyses. These are further divided into 10 subparts. 1A) Models for genetic interactions are formally described including their restrictions. 1B) Boolean adjacency matrices (BAMs) are generated that define the topologies of the models. 1C) Weights are included into the models (Weighted edge matrices, WEMs). 1D) Symmetrical WEMs are removed. 1E) Petri nets are generated. 2A) Initiation of Petri nets with four conditions. 2B) Simulation of Petri nets. 3A) genetic interaction patterns are assigned to the models based on the simulation results from 2B. 2B and 3A are iterated five times. 3B) 'Unstable' models and models with rare patterns are removed. 4) Downstream analyses to suggest generalized mechanisms of genetic interactions.

5.2.2 Defining the topologies of the models and adding weights on edges

The model topologies are defined with BAMs, specifying presence ('1') or absence ('0') of directed edges between nodes from columns to row (Figure 5.2B). All possible four-node models were created: two upstream nodes (interacting gene pair or regulators), R1 and R2, and two downstream nodes (genes), G1 and G2, connected through two to eight directed edges, with three exceptions: 1) no self-edges (i.e., autoregulation), 2) maximum of two incoming edges to any node, and 3) minimum of two outgoing edges from the regulators to the downstream genes.

Between 16 and 65,536 WEMs were created for each BAM, where each 'present' edge ('1' in the BAM) was given a weight of either '-1' (weak inhibition), '-5' (strong inhibition), '1' (weak activation) or '5' (strong activation); an 'absent'

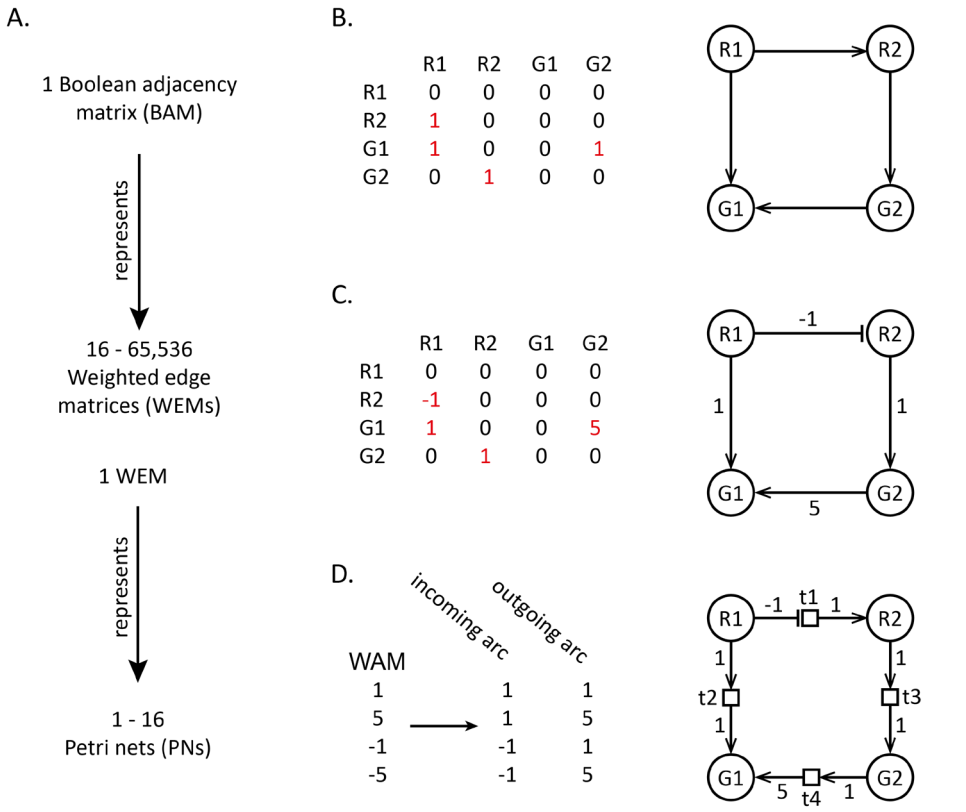


Figure 5.2. Three levels of models were generated, Boolean adjacency matrices (BAMs), weighted edge matrices (WEMs), and Petri net models. A) One BAM represents between 16 and 65,536 WEMs, and one WEM represents between 1 and 16 Petri nets. B) Examples of a BAM, C) one of its possible WEMs, and D) one of its possible Petri nets, showing inversion 'up-no-DOWN', are shown. B) Matrices and their respective network representations are shown for the BAM, and C) the WEM. D) The Petri nets describe the edges by a transition with one incoming arc and one outgoing arc. The conversion rule from a WEM edge weight to the two Petri nets arc weights and the graphical illustration of the Petri net are shown.

edge remained a '0' (Figure 5.2A and 5.2C). Because the two regulators and the two downstream genes are mutually interchangeable, every model has up to three symmetrical counterparts, which were filtered away.

5.2.3 Generation of Petri net models

Between 1 and 16 Petri net models were created for each WEM (Figure 5.2A and 5.2D). Petri nets have two types of nodes, 'places' (genes) and 'transitions' (gene interaction), connected by directed edges called 'arcs'. Here, each node in the WEMs became a transition with one incoming and one outgoing arc. Each place can hold tokens that denote the availability of the corresponding genes. Tokens can move between places through transitions. Arcs can be activating or inhibiting and have weights that denote amount of tokens to consume from or produce to places. The sign of the edge weight in a WEM (edge type) defines if the corresponding Petri net transition is inhibiting or activating and is specified on the arc going from the input place to the transition. The edge weight in a WEM defines if the corresponding Petri net transition is weak or strong and is specified on the arc going from the transition to the output place.

More than one Petri net model can be created from a WEM if at least one node has two incoming edges. A node with two incoming edges can be represented in the Petri net in two ways. A place can be activated by one transition depending on two input places ('AND' logic) or two transitions independent of each other ('OR' logic). Transitions with AND logics were only created when the two edges had the same value in the WEM.

5.2.4 Initiating Petri nets for the four conditions

Each Petri net was simulated with four conditions: 1) wild type (WT), 2) R1 deletion mutant (*R1Δ*), 3) R2 deletion mutant (*R2Δ*), and 4) R1 and R2 deletion mutants (*R1R2Δ*). The gene places, G1 and G2, were initiated with 0 tokens, because the simulation output is measured in these. Deletion mutants were similarly initiated with 0 tokens. WT regulators were initiated with 200 tokens, which in this implementation is equivalent to having unlimited tokens, so the regulation by the regulators stays the same throughout the simulation. Exceptions apply when the two regulators, R1 and R2, are connected to properly reflect the biological nature of genetic interactions in the models as detailed in Figure 5.3A. Furthermore, all incoming edges to deletion mutants were removed to prevent accumulation of token (i.e., activation) in these places (Figure 5.3B).

5.2.5 Petri net simulation rules

Petri nets were simulated in a stepwise manner with maximal parallel execution (Burkhard, 1980; Bonzanni et al., 2009), using rules as set out by Bonzanni et al. (2013) and Bonzanni et al. (2014) (summarized in Figure 5.3C). The models were simulated using a script adapted from Jacobsen et al. (2016). For each of the four conditions of each Petri net we simulated the gene expression levels in G1 and G2 over 50 steps (including the initial step) one time for non-stochastic models, 10 times

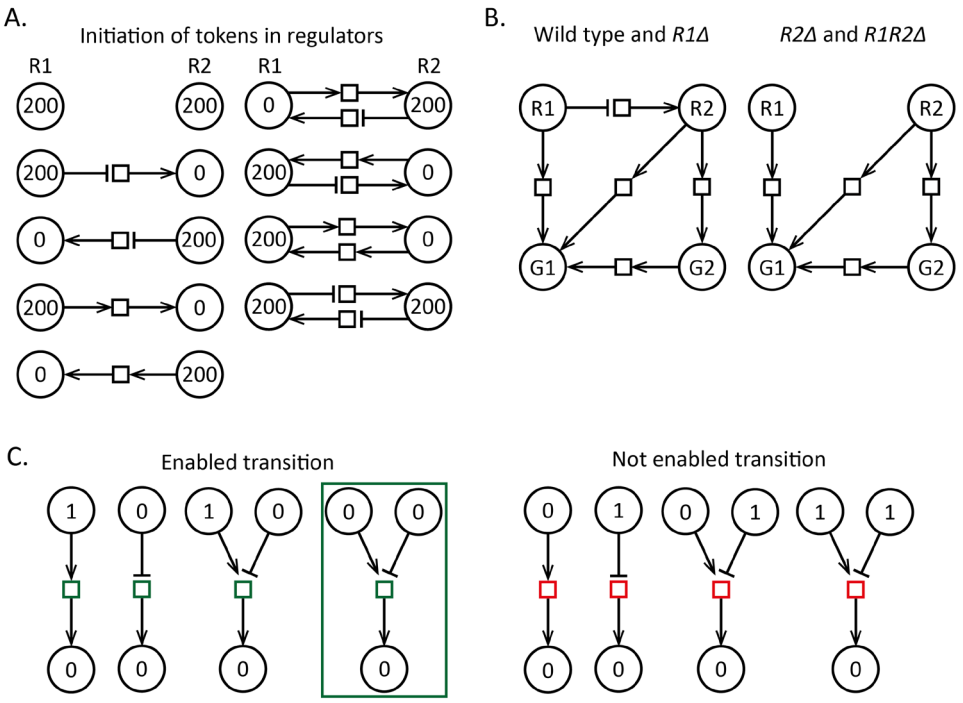


Figure 5.3. Petri net rules. A) The initial token number in R1 and R2 depends on how they are connected. B) Incoming edges to deletion mutants are removed. In the example shown R1 inhibits R2 in the wild type. In *R1Δ* the edges remain, because there are no incoming edges to R1. In *R2Δ*, as well as in *R1R2Δ*, the inhibition from R1 to R2 is removed. C) A transition is enabled for an activating arc if the input place has at least 1 token, and for an inhibiting arc, 0 tokens. For AND logic transitions the same rules apply for both arcs, however with one exception, if an inhibiting arc fulfills the requirements but an activating arc does not, the inhibiting arc over rules.

for stochastic models with one competing connection, and 100 times for stochastic models with more than one competing connection. We calculated the mean of the token levels in the simulation results for each of the genes, G1 and G2, respectively. This whole process (50 steps 1/10/100 times and further mean calculation) was repeated 5 times.

2.2.6 Assigning genetic interaction patterns

For each of the four conditions of each Petri net, the simulation results were used to assign genetic interaction patterns to both genes, G1 and G2, following the criteria set by Sameith et al. (2015). We first calculated the fold change (FC) of the final token number (i.e., gene expression changes) of each downstream gene in each mutant compared to WT: $FC = \log_2(\text{mutant}/\text{WT})$, where mutant is the final token count in the respective gene after simulation of a mutant condition, and WT is the final token count in the respective gene after simulation of the WT condition. The FC was used to calculate the genetic interaction (GI) score for each downstream

gene in each model: GI = FC(*R1R2Δ*) - (FC(*R1Δ*) + FC(*R2Δ*)), where FC(x) denotes the FC in the corresponding mutants. The GI scores and FCs were used to map genetic interaction patterns to the genes in Petri nets. The mappings of the observed phenotypes (gene expression) to genetic interaction patterns from Sameith et al. (2015) were used: the gene expression changes (up-, no-, or down regulation) in the mutants compared to WT, and the sign of the GI score correspond to a specific pattern. The threshold for defining up and downregulation, as well as positive and negative genetic interaction, was $\log_2(1.7)$ and $-\log_2(1.7)$, respectively.

The threshold of $\pm\log_2(1.7)$ was established previously (Amini et al., 2018): between $-\log_2(1.7)$ and $\log_2(1.7)$ the gene was assigned 'non-interacting'. Some models with a GI score above $\log_2(1.7)$ were assigned to 'miscellaneous' and 'miscellaneous simulation' patterns, because they did not correspond to one of the six genetic interaction patterns. Such patterns were found to be rare in Sameith et al. (2015) or in the simulation results.

When both genes in a model are consistent in the pattern assignment for each of the five simulations the model is considered 'stable', however, when at least one gene is assigned to different patterns for one or several simulations, the model is deemed 'unstable'. To avoid uncertainty in the simulation result unstable models were removed based on the assumption that only stable models are meaningful in real gene interaction networks.

5.2.7 Investigating the importance of network complexity and edge strengths in genetic interaction models

We calculated the number of Petri nets per complexity level for each pattern and frequencies of edge strengths per complexity level. These numbers were then normalized by the total number of Petri nets, N, within the same complexity level, k, as: $N_{k,\,pattern}/N_{k,\,total}$.

5.2.8 Generalised mechanisms of inversion and suppression

To allow comparison between the model topologies we considered symmetry around G1 and G2, where G1 is the node under consideration, and G2 is the 'intermediate' node, which also may show genetic interaction. So, G1 and G2 were exchanged for all models where G2 is the node under consideration. Further, we also considered symmetry around R1 and R2 to distinguish the effect of their respective single mutations. Moreover, each genetic interaction pattern consists of two or four subpatterns, depending on the direction of the mutations; these are illustrated in S5.1 Figure.

For each subpattern we expect that most models with low complexity levels describe models with higher complexity levels. Thus, we can describe each subpattern with a set of unique topologies, 'prototypes', that may explain higher complexity level models, while retaining the same behavior (genetic interaction pattern). A 'prototype', thus, generalizes all its inclusive Petri nets. For each prototype of inversion ('up-no-DOWN' and 'down-no-UP') and suppression ('up-no-NO' and 'down-no-NO') we determined and counted all topologies and their

corresponding Petri nets (S5.2-S5.5 Figures). We normalized the number of Petri nets for each prototype by the number of models in the subpattern, with complexity levels starting from the complexity level of the prototype.

We described generalized mechanisms for the five most abundant prototypes for inversion 'up-no-DOWN' and 'down-no-UP' and suppression 'up-no-NO' and 'down-no-NO' by also taking the edge weight frequencies into consideration (S5.6-S5.9 Figures). If the edge is activating or inhibiting in more than 75% of the Petri nets, this is indicated on the edges (arrow for activation and perpendicular bar for inhibition). By default a circle is used. Likewise, weight (-1, -5, 1 or 5) is shown on an edge if it appears in more than 75% of the Petri nets, (bold edge for strong). Finally, strength (weak: -1/1, or strong: -5/5 with bold edge) is shown on an edge if it appears in more than 75% of the Petri nets, where type and weight are not dominant. Additional edges are considered relevant for the mechanism, if a particular non-zero edge weight, type, or strength, occurs in more than 50% of the Petri nets for that prototype.

5.3 Results

5.3.1 Definition and exhaustive generation of genetic interaction models

In this study, we created a framework for exhaustive modeling of genetic interaction patterns using Petri nets (Figure 5.1). We defined four-node models connected through two to eight directed edges with weights of 1, 5, -1, or -5 (weak, strong activation, weak, strong inhibition). Two incoming edges are implemented either dependent (AND logics) or independent (OR logics) of each other, in seperate models. A model has two upstream regulators, R1 and R2, that control the expression of two downstream genes, G1 and G2. Via and edge, each gene mediates regulation of another gene. Genetic interaction between the regulators may be detected in either one of the downstream genes, or in both. Filtering included three restrictions for creating BAMs (yielding 728) and removal of symmetrical WEMs (yielding 2,323,936; see Methods), resulting in a total of 9,172,034 Petri net models.

5.3.2 Petri net simulation of genetic interaction patterns

All Petri nets were simulated with four conditions: WT, two single and one double deletion mutant (see Methods). A GI score was calculated for each gene to map them to possible genetic interaction patterns, using the procedure described by Sameith et al. (2015). Models for which the assigned genetic interaction pattern in at least one of the genes was inconsistent between five simulations were labeled 'unstable' and discarded (212,169, 2.3%). 1,546,825 (17.3%) of the stable models were discarded because they were assigned to miscellaneous patterns (683,831, 7.6%) as defined by Sameith et al. (2015), to additional miscellaneous patterns only observed in the Petri net simulations (838,755, 9.4%), or both (24,239, 0.3%). 5,144,630 (57.4%) were assigned to one or two genetic interaction patterns, the remaining 2,268,410

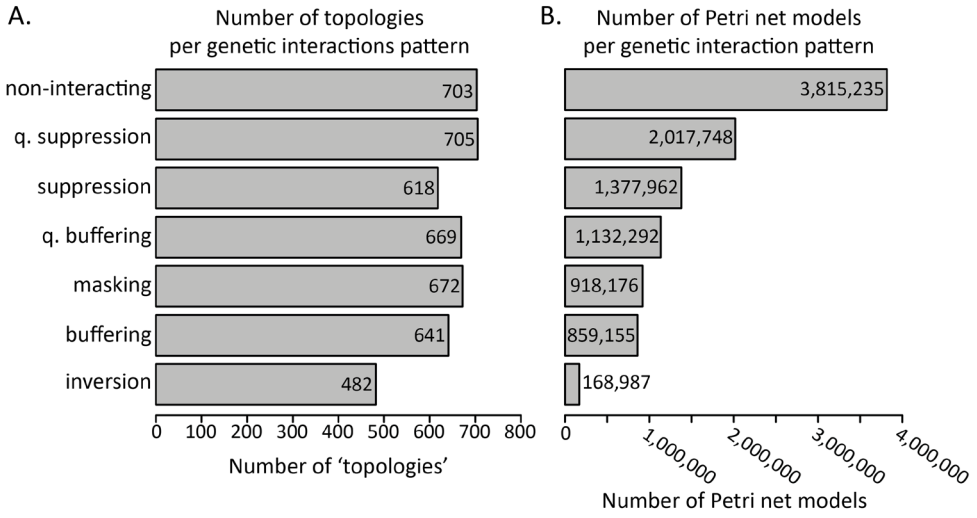


Figure 5.4. Number of topologies and Petri net models for each genetic interaction pattern: quantitative suppression, suppression, quantitative buffering, masking, buffering, and inversion, and non-interacting. A) Number of topologies per genetic interaction pattern out of the total 728 topologies constituting the stable Petri nets. B) Number of Petri nets per genetic interaction pattern out of the total 8,959,865 stable Petri nets. Petri nets assigned to two genetic interaction patterns are counted twice (for G1 and G2). In this figure quantitative is abbreviated as 'q.'.

(25.3%) were non-interacting in both genes. The number of topologies and Petri nets for each of the genetic interaction patterns and those showing no interaction are shown in Figure 5.4. Inversion is observed in fewest topologies and Petri nets (482 and 168,987) and quantitative suppression in most (705 and 2,017,748). Each genetic interaction pattern constitutes a large fraction of the total 728 topologies, thus there does not seem to be an obvious connection between particular topologies and genetic interaction patterns.

5.3.3 Network complexity and edge strength is important for genetic interaction patterns

We investigated the relation between edge weights and assigned genetic interaction pattern by counting the number of 'strong' edges (weight +5 or -5) in each Petri net. Figure 5.5 shows normalized frequencies of Petri nets and strong edges per complexity level (number of edges) for inversion, masking, buffering and suppression. Independent of edge weight the overall distribution of Petri nets is skewed towards higher complexity levels for each genetic interaction pattern, but with different shapes. The distribution increases linearly for inversion and masking, has two peaks for buffering with peaks at complexity level 2 and 8, and peaks at complexity level 6 for suppression. Note that there are no two-edge models for suppression and inversion. Quantitative buffering and quantitative suppression show similar distributions as their non-quantitative counterparts, and models showing no interaction show higher overall frequencies (just above 0.4) than the models with

assigned genetic interaction patterns (S5.10 Figure). For non-interacting models, complexity level 2 is substantially more likely, indicating that the low complexity most likely leads to non-interacting behavior.

The frequency of strong edges is visualized by the colors of the bars in Figure 5.5, where red means no strong edge and the other rainbow colors one to eight strong edges in a model. The maximum number of possible strong edges equals the total

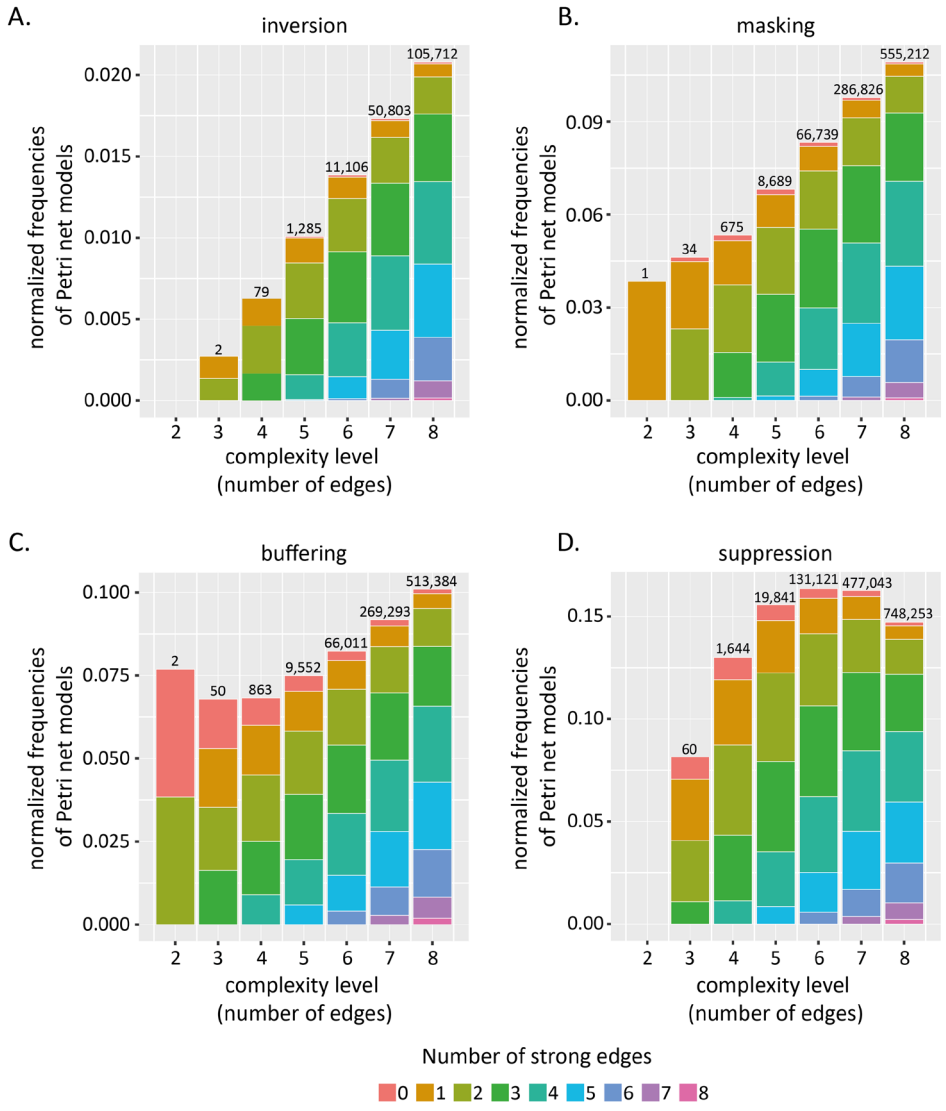


Figure 5.5. Normalized frequencies of Petri net models and 'strong' edges per complexity level (number of edges) per genetic interaction pattern. The distributions are shown for A) inversion, B) masking, C) buffering, and D) suppression. The number above each bar represents the actual number of Petri nets for that complexity. The colors represent the number of strong edges (weight +5/-5) in the Petri net.

number of edges, hence the number of colors per bar increases with the complexity level. Further, we expect a binomial distribution for these frequencies, as models with no strong or no weak edges can only be described in one way (all edges are strong or all are weak), whereas models with both strong and weak edges can be described in multiple ways. Thus, intermediate numbers of strong edges, relative to the complexity level, are expected to occur most often. Only 993 out of the 168,987 inversion models (0.59%) have no strong edge. Likewise only 1,243 (0.73%) of the models have no weak edge. Note that topologically identical models with all weak edges or all strong edges have identical behavior, only all token levels will differ five-fold; while the relative levels determining genetic interaction patterns stay the same. The number of models with all weak/strong edges is 1.32% and this is much smaller than the 2.8% expected from the corresponding binomial distributions for all possible networks with two to eight edges, which demonstrates that a combination of both weak and strong edges is crucial for inversion. This effect is also important for masking, but slightly less so, as 0.75%, and 0.93% masking models are observed with no strong edges and no weak edges, respectively. For buffering and the other patterns we observe the binomially expected fraction of around 1.4% of models with only weak or only strong edges for each complexity level.

In Figure 5.4A, we see that each genetic interaction pattern constitutes a large fraction of the total 728 topologies, thus there is no obvious connection between a particular topology and a particular genetic pattern. However, in Figure 5.5 we see that the distributions of Petri nets across complexity levels, and of edge strengths within complexity levels differ between the genetic interaction patterns. Thus, since network complexity is a topological feature, topology does play a role in the underlying mechanisms of genetic interaction patterns, in combination with other factors as we e.g. observed for inversion and masking, where both weak and strong. To sum up, the combination of strong and weak edges may be considered as a characteristic property of inversion and masking patterns.

5.3.4 'Prototypes' enable description of generalized mechanisms for genetic interaction patterns

To better understand the regulatory mechanisms of genetic interaction patterns in the simulation results, we would like to derive some generalized models based on the large variety of models that we see associated with each genetic interaction pattern. In order to do that, we also first must separate the genetic interaction patterns into subpatterns based on their behavior. For example, inversion, where the double mutant has the opposite effect of one of the single mutants, comes in four varieties: either one, or both, of the single mutants may show an effect, and this may be either (one or both) positive, or (one or both) negative. We write these as 'up-no-DOWN', 'up-up-DOWN', 'down-no-UP' and 'down-down-UP', respectively (see S5.1 Figure for a complete overview). We introduce the concept of a 'prototype', which is the topology with the lowest complexity that is found in included in (higher complexity) topologies of Petri nets of the same genetic interaction subpattern. Here, we will

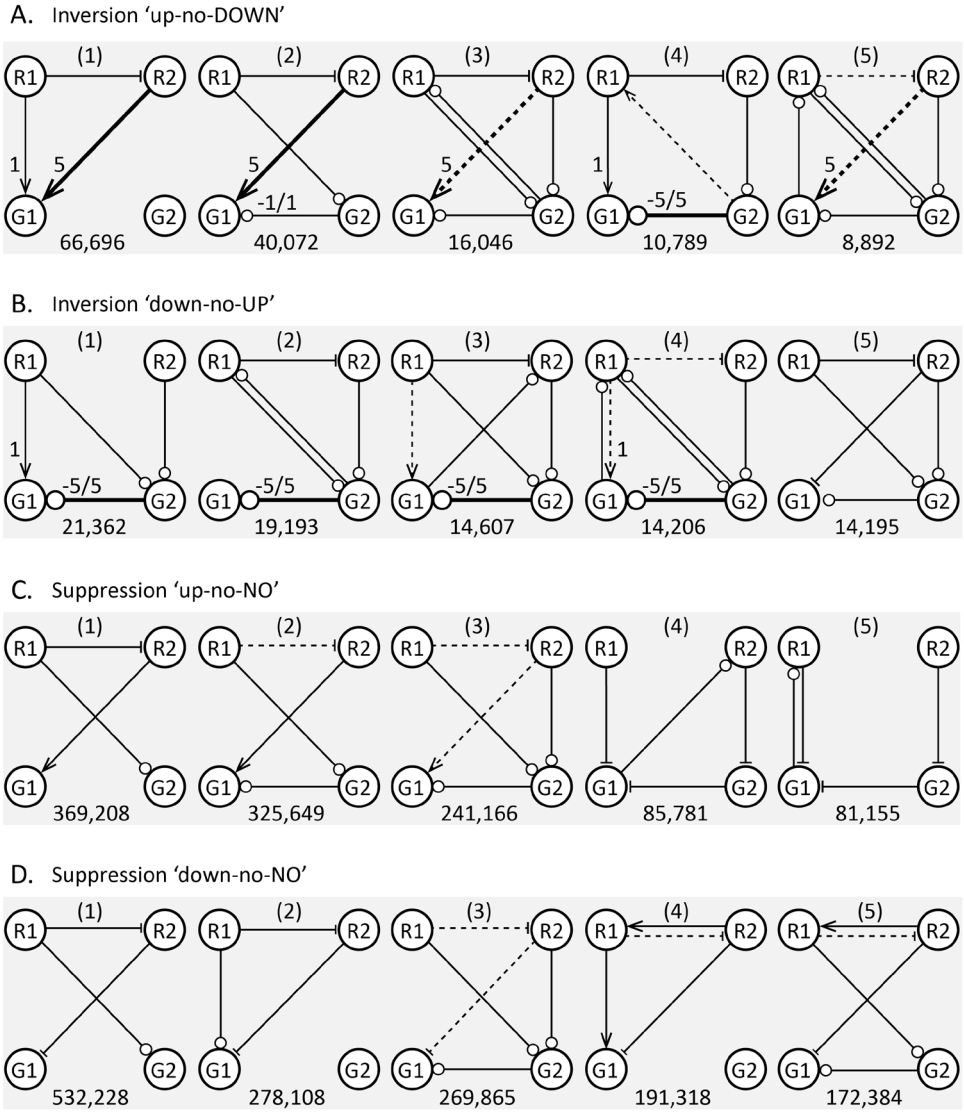


Figure 5.6. Generalized mechanisms for the genetic interaction patterns, inversion and suppression. Five mechanisms are described by the five most abundant prototypes with frequent edges and edge weights for A) inversion 'up-no-DOWN', B) inversion 'down-no-UP', C) suppression 'up-no-NO', and D) suppression 'down-no-NO' are shown. G1 is always the one showing inversion (A,B) or suppression (C,D). Edges with an empty circle mean no clear signal on activation or inhibition. Edge weights, type and strenght are shown if these are seen in the majority of the Petri nets (>75%). Dashed edges are added to the prototype, if they are present in more than 50% of the Petri nets of that prototype. The numbers of Petri nets per prototype is shown below each graph.

focus on the five most abundant prototypes for the two most frequent subpatterns of inversion and suppression. 'up-no-DOWN' and 'down-no-UP' represent 96.67% of all inversion models, and 'up-no-NO' and 'down-no-NO' represent 99.73% of all suppression models. For these, we analyzed edge weight frequencies to include information about the type (activating or inhibiting) and strength (1 or 5) of the edges occurring in the majority of the Petri nets (see Methods for details) in order to describe their generalized mechanisms (see Figure 5.6). Note that, through the way we set up the analysis, in these models G1 is always the one that shows the inversion (Figure 5.6A and 5.B) or suppression (Figure 5.6C and 5.6D) pattern.

5.3.5 Generalized mechanisms for inversion

We identified 19 prototypes for inversion 'up-no-DOWN', which encapsulate all 120,255 Petri nets for that subpattern, and 45 for inversion 'down-no-UP' encompassing all 49,216 models (S5.2 and S5.3 Figures). The five most abundant prototypes for each inversion subpattern are shown in Figure 5.6A-B, respectively.

The simplest mechanism for inversion 'up-no-DOWN' ('1' in Figure 5.6A) has the following three components: R1 weakly activates G1, R2 strongly activates G1, and R1 inhibits R2. Thus, R1 blocks the strong activation of R2 on G1. Therefore, there is 'no' difference between WT and *R2Δ*. However, in *R1Δ*, R2 is 'activated', and G1 goes 'up', while in *R1R2Δ* G1 goes 'down'. Mechanisms 2 and 4 (Figure 5.6A) have the same three components, but the regulation from R1 or R2 to G1 is indirect via G2. In mechanisms 3 and 5 (Figure 5.6A) both are indirect: the regulations from R1 and R2 to G1 both go via G2.

For inversion 'down-no-UP' the simplest mechanism ('1' in Figure 5.6B) can be explained with: R1 weakly activates G1, and R1 and R2 strongly inhibit G1 indirectly via G2. Thus both R1 and R2 contribute to the inhibition of G1; hence there is 'no' difference between WT and *R2Δ*, and in *R1Δ* G1 goes 'down', while in *R1R2Δ* G1 goes 'up'. The weak activation from R1 to G1 is also seen in mechanisms 3 and 4. Different from mechanism 1, R1 inhibits R2 in mechanisms 2-5. Thus, there is 'no' difference between WT and *R2Δ*, where R2 is either blocked by R1 or knocked out leading to a weak activation in G1 via G2. In *R1Δ*, R2 is 'activated' and G1 goes 'down'. In *R1R2Δ*, G1 goes 'up' because of activation from both R1 and R2.

5.3.6 Generalized mechanisms for suppression

We identified 19 prototypes for suppression 'up-no-NO', which encapsulate all 609,778 Petri nets for that subpattern, and 22 prototypes for suppression 'down-no-NO' encompassing all 882,656 models (S5.4 and S5.5 Figures). Figure 5.6C and 5.6D show the five most abundant prototypes for suppression with added frequent edges and dominant edge weights.

The simplest mechanism for suppression 'up-no-NO' ('1' in Figure 5.6C) has three components: R1 inhibits R2, R2 activates G1, and an interaction from R1 to G2. The first two interactions add up to an indirect inhibition of G1 by R1, via R2. This pattern re-occurs in mechanisms 2 and 3 (Figure 5.6C). In mechanisms

4 and 5, R1 directly inhibits G1, while now R2 activates G1 indirecty by a double inhibition via G2. Because R1 directly and indirectly inhibits G1, there is 'no' difference between WT and *R2Δ*. In *R1Δ* this inhibition is removed and G1 goes 'up'. In *R1R2Δ* G1 is not activated and has 'no' difference between WT.

The mechanisms for suppression 'down-no-NO' are quite comparable to 'up-no-NO' mechanisms 1-3, with the difference that R2 inhibits G1 (instead of activates), yielding an indirect activation of G1 by R1 through a double inhibition via R2, and in mechanisms 3 and 5 also via G2. Thus, there is 'no' difference between WT and *R2Δ* and *R1R2Δ*, where the inhibition of R1 on R2 leads to an activation of G1. This is blocked, however, in *R1Δ*, where G1 goes 'down'.

In many of the mechanisms in Figure 5.6 for both inversion and suppression, other interactions occur that we did not describe in our interpretations of the mechanisms. However, these are not easily characterized and further investigation is required to identify their roles in the GI patterns of suppression and inversion.

5.4 Discussion

In this study, we have addressed the regulatory complexity of genetic interactions observed in four-node networks with a computational approach. The exhaustive modeling framework presented in this study, was also applied in (Amini et al., 2018), and to the best of our knowledge is the first that investigates all possible underlying mechanisms for a set of genetic interaction patterns, where previous approaches have focused on observed genetic interactions in specific gene-pairs (Pirkl et al., 2017). For the genetic interaction patterns observed in *S. cerevisiae* by Sameith et al. (2015), we observed that network complexity and edges strengths play a role in their mechanisms. By systematically studying the model properties we derived putative generalized mechanisms for inversion and suppression, two poorly characterized genetic interaction patterns.

An entirely unbiased exhaustive approach with four fully connected nodes and five edge weights (parameters) would produce 5^{16} = 152,587,890,625 models, which renders unfeasible any kind of comprehensive modeling strategy. To make an exhaustive approach feasible we limited ourselves in the definition of the model topologies and parameters. To be able to do that efficiently, we generated the models on three levels with filtering steps in between. 4,096 Boolean adjacency matrices (BAMs) were generated with three restrictions producing 2,323,936 weighted edge matrices (WEMs) after filtering symmetrical models. As results, we arrived at slightly over 9 million biologically relevant Petri net models for simulation and analysis, which is a manageable number, and incidentally only a small fraction of the initial 152 billion.

We showed that each genetic interaction pattern constitutes a large fraction of the total 728 topologies (see Figure 5.4A), thus there is no obvious connection between a particular topology and a particular genetic interaction pattern. However, in Figure 5.5 we see that the distributions of Petri nets across complexity levels, and of edge strengths within complexity levels differ between the genetic interaction

patterns. Thus, since network complexity is a topological feature, topology does play a role in the underlying mechanisms of genetic interaction patterns, in combination with other factors as we e.g. observed for inversion and masking, where both weak and strong edges are important.

To describe plausible generalized mechanisms for genetic interaction patterns, we defined the concept of a 'prototype' that describes a set of Petri net models with shared topological features and equivalent genetic interaction subpattern. In a study by (Videla et al., 2017) using prior-knowledge networks (PKN) and experimental data, families of logical networks were learned from a given topology; models in one family characterize the response expected from experiments equally well. A family of models in their framework is comparable to our definition of prototypes. Further, combining prototypes with frequent edges and dominant edge weights enabled us to identify generalized mechanisms for inversion and suppression (see Figure 5.6). In these generalized mechanisms the combination of strong and weak edges, as well as activating and inhibiting edges, were crucial properties. Thus, a quantitative difference characterizes these patterns. This, which is in agreement with a study by Pirkl et al. (2017) where they inferred the networks for particular gene sets in van Wageningen et al. (2010) and Sameith et al. (2015), from which they concluded that quantitative properties are needed to reproduce all patterns in the dataset. In addition to that more general conclusion, the use of the prototype analysis also allowed us to investigate the correlated occurrence of certain edges, edge types and edge weights that in combination leads to particular behavior of the network.

Our results show that genetic interaction patterns can be explained by a vast number of Petri nets. However, we have no gold standard to assess the relevance of our simulation results, as a complete solution space cannot be observed with current experimental technology. The experimental data on gene specific transcription factors (GSTFs) in *S. cerevisiae* by Sameith et al. (2015) favors buffering and inversion, while in our simulation results suppression and quantitative suppression occur most frequently (see Figure 5.4). Some of these differences no doubt arise from biases in our modeling approach. However, they may also be explained by functional properties of GSTFs, which might yield particular selection pressure for certain genetic interaction patterns (Amini et al., 2018). In this sense, one may also interpret our overall set of Petri net models as an elaborate and well-informed null-model of which distributions of network properties and genetic interaction patterns to expect in the absence of (natural) selection. Milo et al. (2002) identified two significantly recurring networks motifs of three- and four-nodes, respectively, in gene interaction networks in yeast. In future studies this type of knowledge can be used to select biologically plausible models.

In summary, this study shows how to model genetic interaction patterns in an unbiased way and offers profound insights into the nature of these genetic interactions by revealing key properties and suggesting potential underlying mechanisms. Our simulation results can be used for subsequent analysis, e.g. to

further investigate the importance of the parameters and their combinations for patterns, and to support a deeper investigation of mechanisms for the other genetic interaction patterns, such as buffering and masking. And, finally, these results may provide a source of inspiration to guide future experimental studies for further improving our understanding of molecular interactions between specific genes.

References

Amini S, Jacobsen A, Ivanova O, Lijnzaad P, Heringa J, Holstege FCP, et al. The ability of transcription factors to differentially regulate gene expression is a crucial component of the mechanism underlying inversion, a frequently observed genetic interaction pattern. *bioRxiv* 2018; doi:10.1101/449520. **Chapter 4 of this thesis.**

Babu M, Arnold R, Bundalovic-Torma C, Gagarinova A, Wong KS, Kumar A, et al. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in Escherichia coli. *PLoS Genet*. 2014, 10, e1004120.

Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. *Bioinformatics*. 2009, 25, 2049-56.

Bonzanni N, Garg A, Feenstra KA, Schutte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013, 29, i80-8.

Bonzanni N, Feenstra KA, Fokkink W, Heringa J. Petri Nets Are a Biologist's Best Friend. Formal Methods in Macro-Biology. Springer, Cham; 2014. pp. 102-16.

Burkhard H-D. On Priorities of Parallelism. Logic of Programs and Their Applications, Proceedings. London, UK, UK: Springer-Verlag; 1983. pp. 86-97.

Chaouiya C. Petri net modelling of biological networks. *Brief Bioinform*. 2007, 8, 210-9.

Costanzo M, Baryshnikova A, Myers CL, Andrews B, Boone C. Charting the genetic interaction map of a cell. *Current Opinion in Biotechnology*. 2011, 22, 66-74.

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science*. 2016, 353.

Haydarlou R, Jacobsen A, Bonzanni N, Feenstra KA, Abeln S, Heringa J. BioASF: a framework for automatically generating executable pathway models specified in BioPAX. *Bioinformatics*. 2016, 32, i60-i9. **Chapter 6 of this thesis.**

Jacobsen A, Heijmans N, Verkaar F, Smit MJ, Heringa J, van Amerongen R, et al. Construction and Experimental Validation of a Petri Net Model of Wnt/beta-Catenin Signaling. *PLoS One*. 2016, 11, e0155743. **Chapter 2 of this thesis.**

Lagator M, Paixao T, Barton NH, Bollback JP, Guet CC. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *Elife*. 2017, 6.

Large EE, Padmanabhan R, Watkins KL, Campbell RF, Xu W, McGrath PT. Modeling of a negative feedback mechanism explains antagonistic pleiotropy in reproduction in domesticated Caenorhabditis elegans strains. *PLoS Genet*. 2017, 13, e1006769.

Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet*. 2006, 38, 896-903.

Mackay TFC, Moore JH. Why epistasis is important for tackling complex human disease genetics. *Genome Medicine*. 2014, 6.

Mayo M, Beretta L. Modelling epistasis in genetic disease using Petri nets, evolutionary computation and frequent itemset mining. *Expert Systems with Applications*. 2011, 38, 4006-13.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. *Science*. 2002, 298, 824-7.

Pe'er D. Bayesian network analysis of signaling networks: a primer. *Sci STKE*. 2005, pl4.

Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008, 9, 855-67.

Pirkl M, Diekmann M, van der Wees M, Beerenwinkel N, Frohlich H, Markowetz F. Inferring modulators of genetic interactions with epistatic nested effects models. *PLoS Comput Biol*. 2017, 13, e1005496.

Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, et al. Signaling and Circuitry of Multiple MAPK Pathways Revealed by a Matrix of Global Gene Expression Profiles. *Science*. 2000, 287, 873-80.

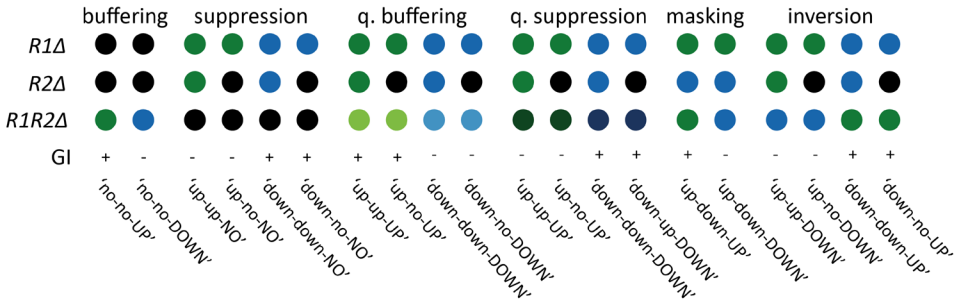
Sameith K, Amini S, Groot Koerkamp MJ, van Leenen D, Brok M, Brabers N, et al. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. *BMC Biol*. 2015, 13, 112.

Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. *Science*. 2004, 303, 808-13.

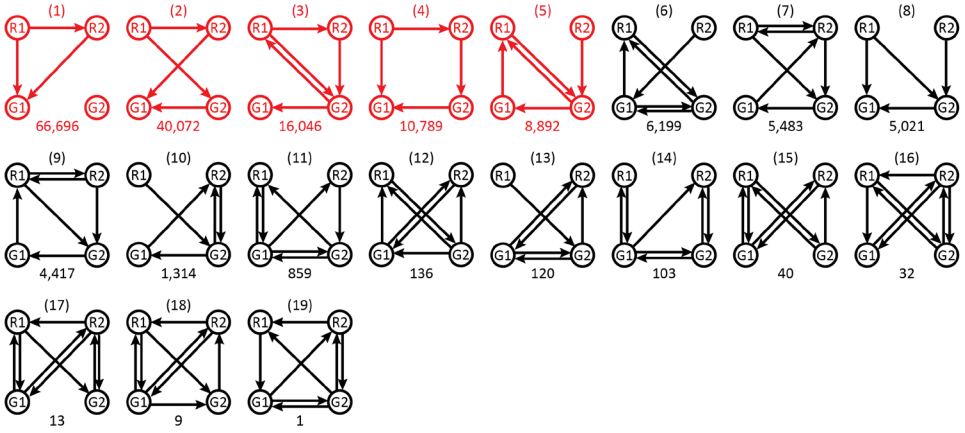
van Leeuwen J, Pons C, Boone C, Andrews BJ. Mechanisms of suppression: The wiring of genetic resilience. Prospects & Overviews. 2017.

van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell*. 2010, 143, 991-1004.

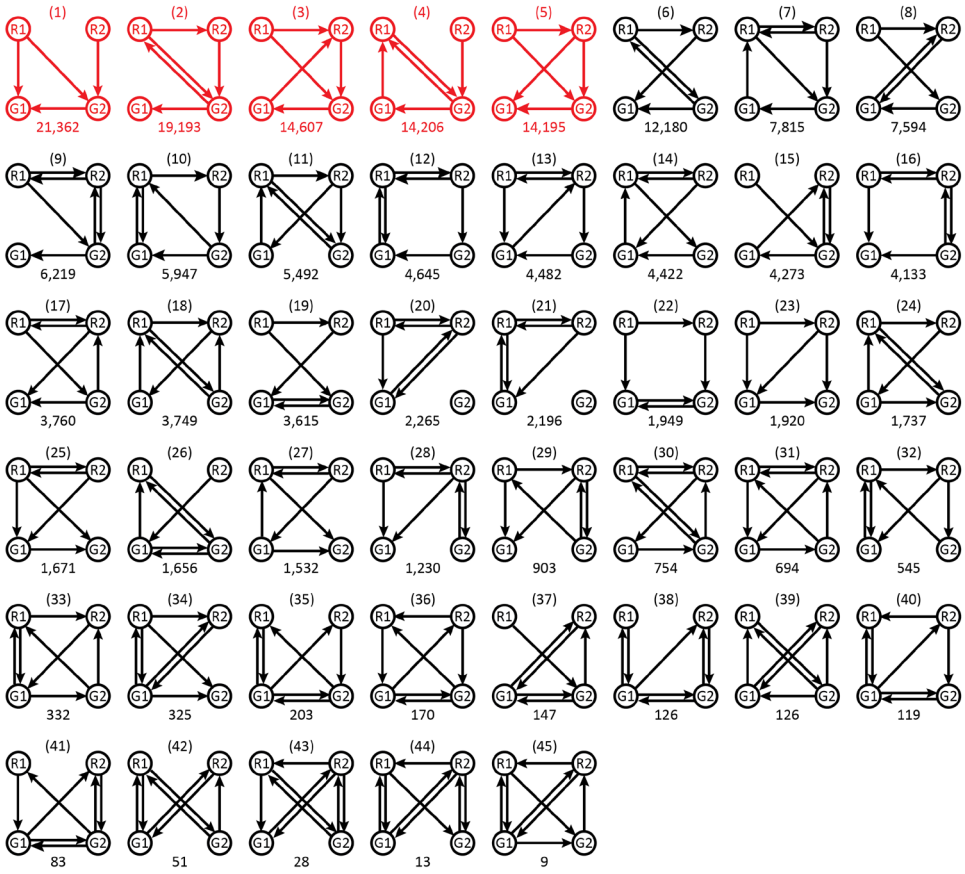
Videla S, Saez-Rodriguez J, Guziolowski C, Siegel A. caspo: a toolbox for automated reasoning on the response of logical signaling networks families. *Bioinformatics*. 2017, 33, 947-50.



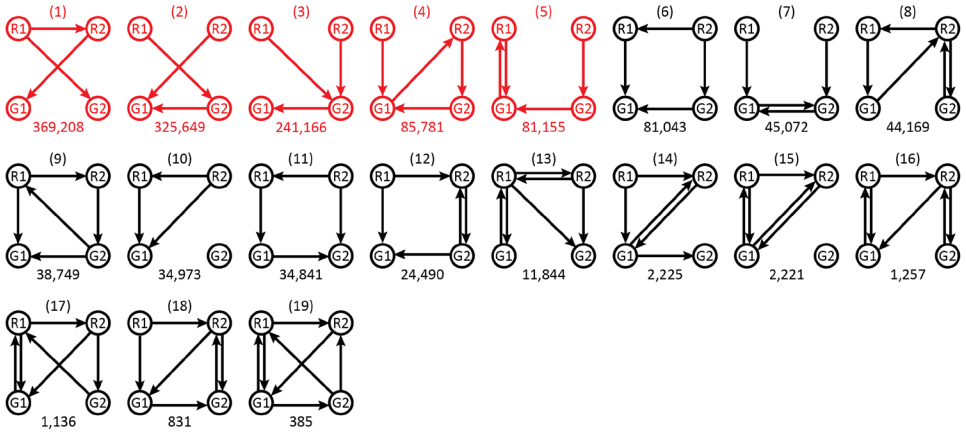
S5.1 Figure. Phenotypes for wild type, single and double mutants for the genetic interaction patterns: buffering, suppression, quantitative buffering, quantitative suppression, masking and inversion (adapted from Sameith et al., 2015). In this figure quantitative is abbreviated as 'q.'.



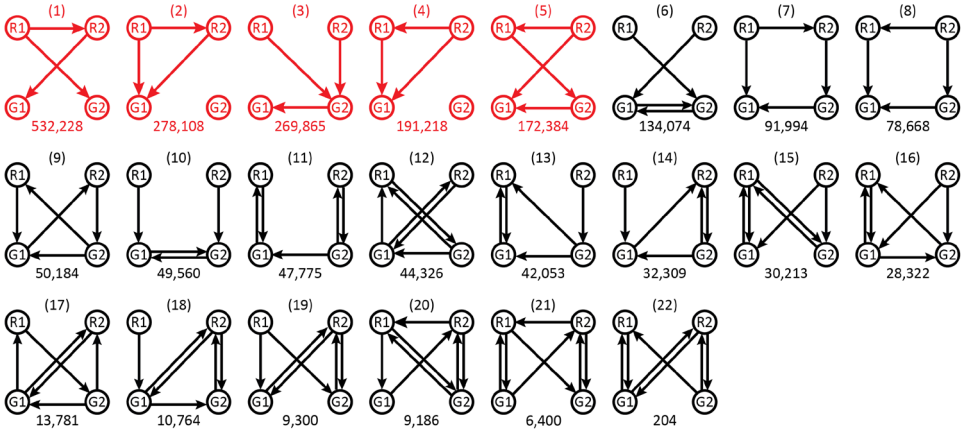
S5.2 Figure. Prototypes for the genetic interaction subpattern: inversion 'up-no-DOWN'. The five most frequent prototypes are shown in red. The arrow on the edges indicates directionality. The number below each prototype represents the number of Petri net models associated with this prototype.



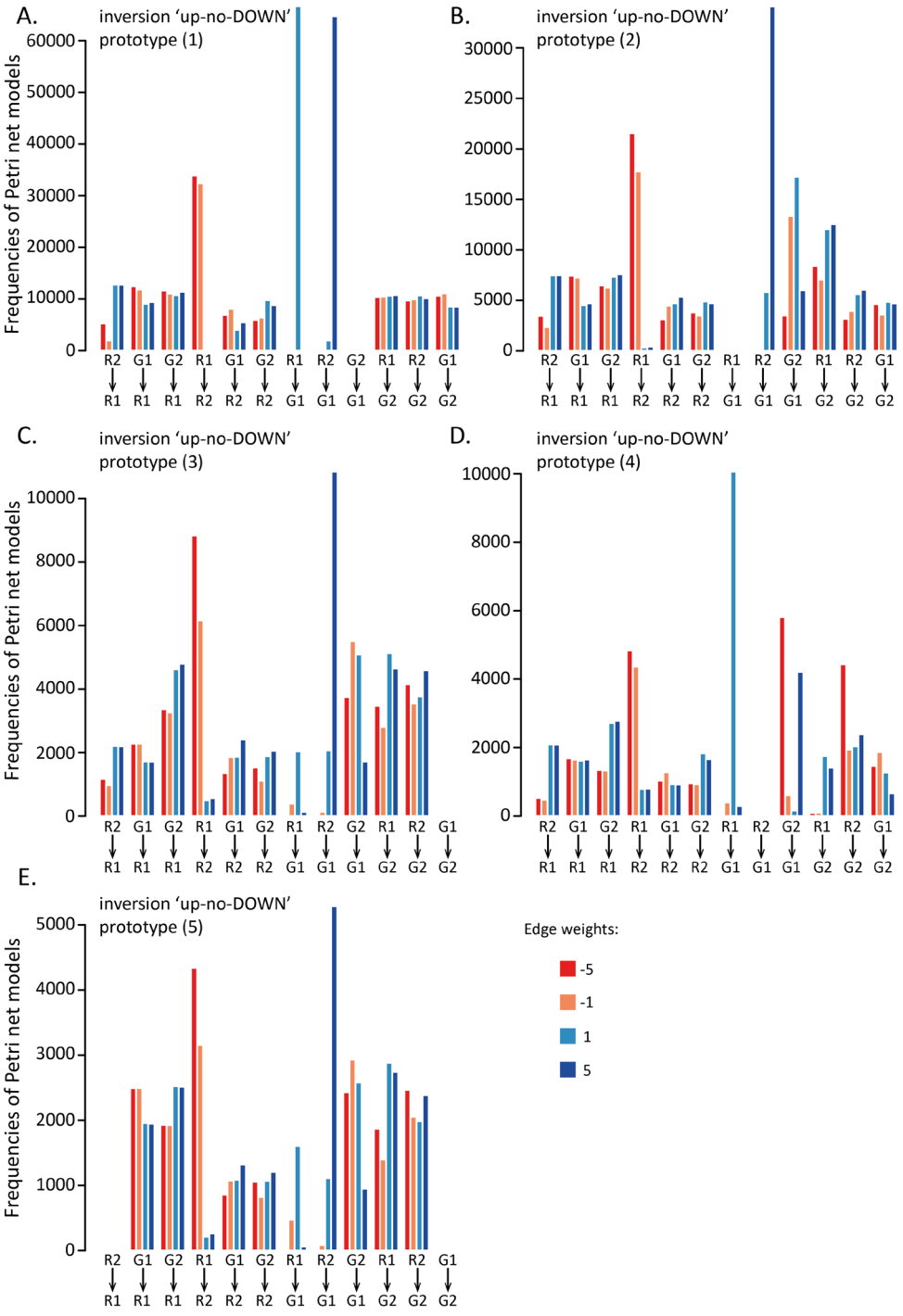
S5.3 Figure. Prototypes for the genetic interaction subpattern: inversion 'down-no-UP'. The five most frequent prototypes are shown in red. The arrow on the edges indicates directionality. The number below each prototype represents the number of PN models associated with this prototype.



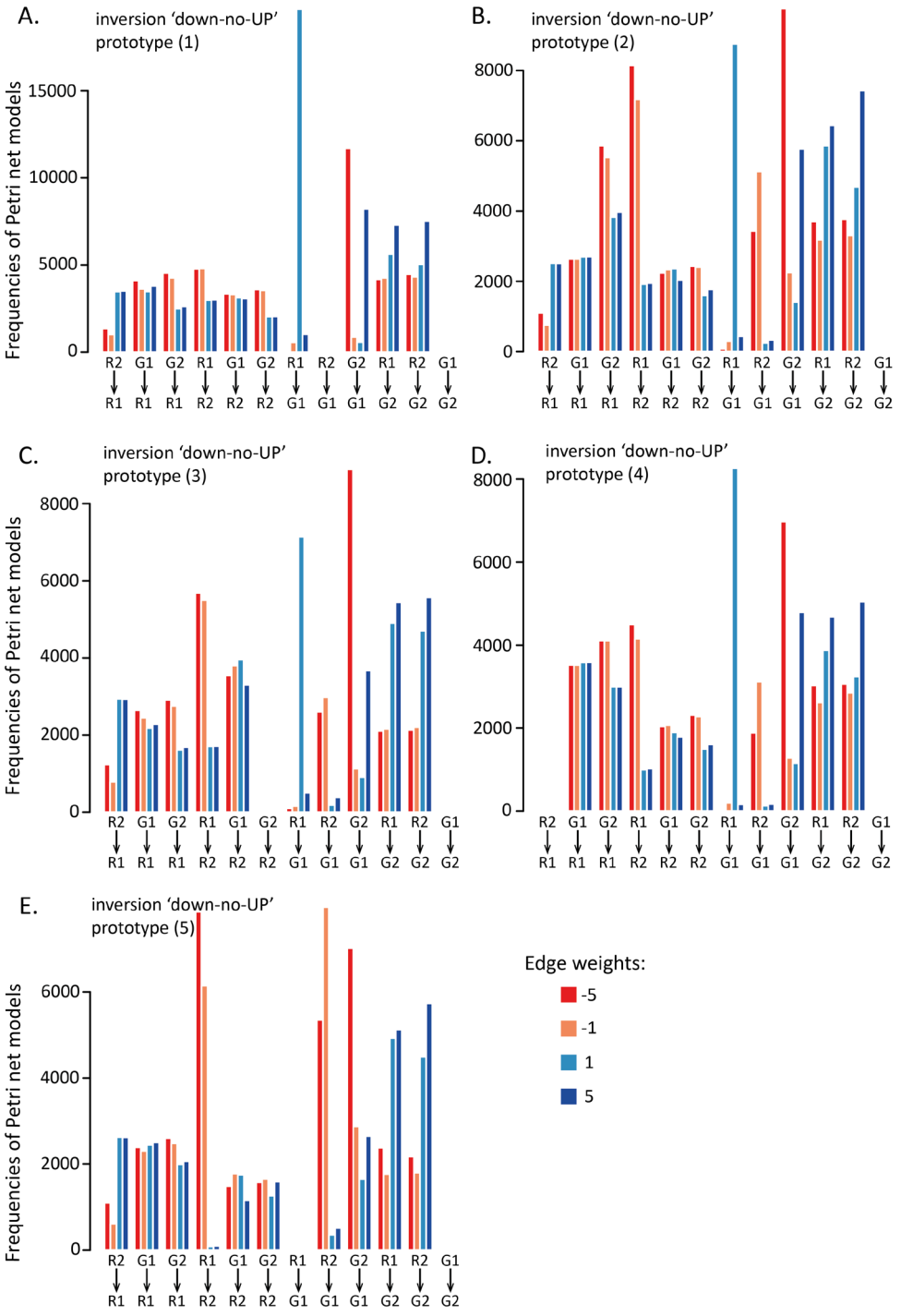
S5.4 Figure. Prototypes for the genetic interaction subpattern: suppression 'up-no-NO'. The five most frequent prototypes are shown in red. The arrow on the edges indicates directionality. The number below each prototype represents the number of PN models associated with this prototype.



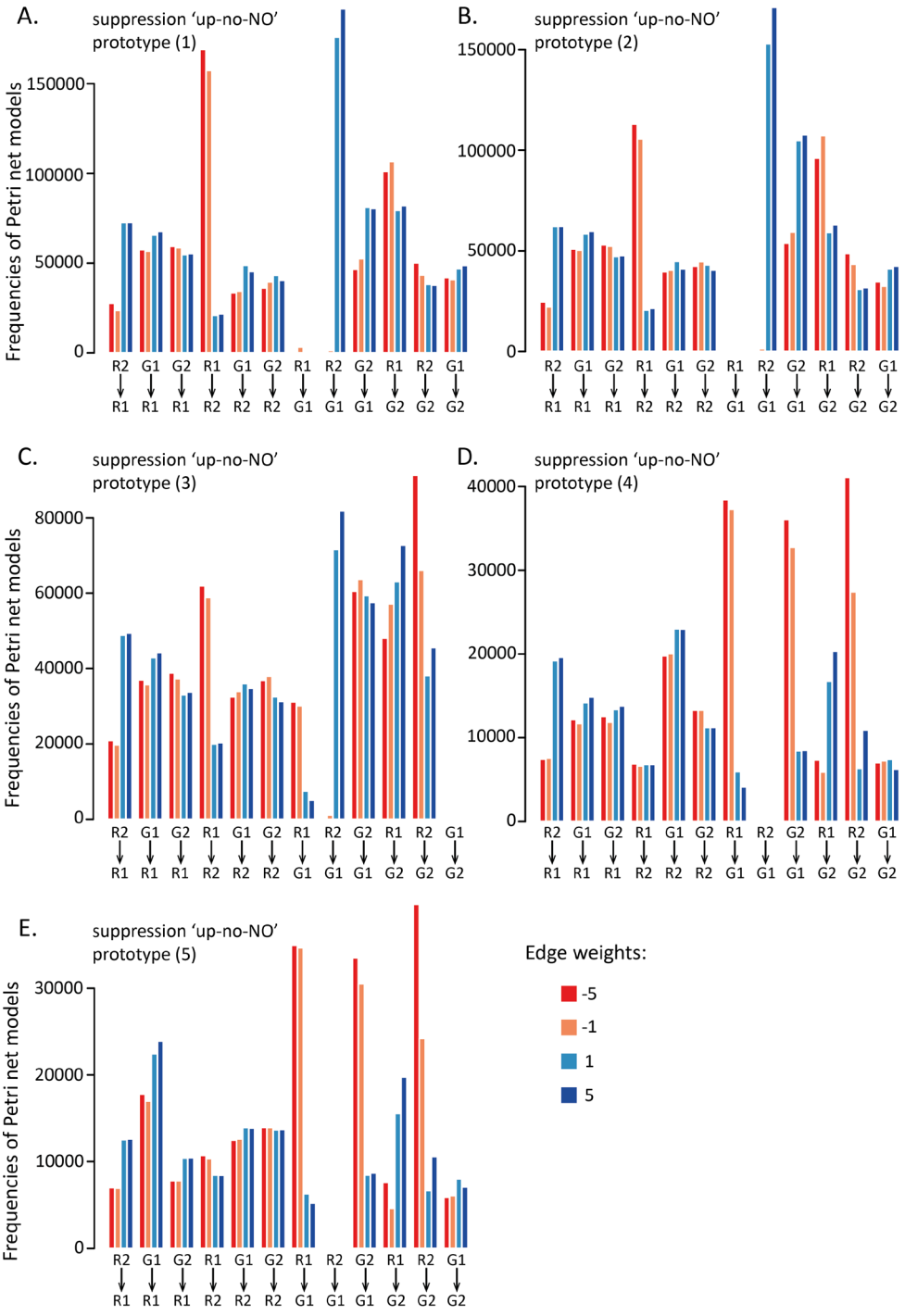
S5.5 Figure. Prototypes for the genetic interaction subpattern: suppression 'down-no-NO'. The five most frequent prototypes are shown in red. The arrow on the edges indicates directionality. The number below each prototype represents the number of PN models associated with this prototype.



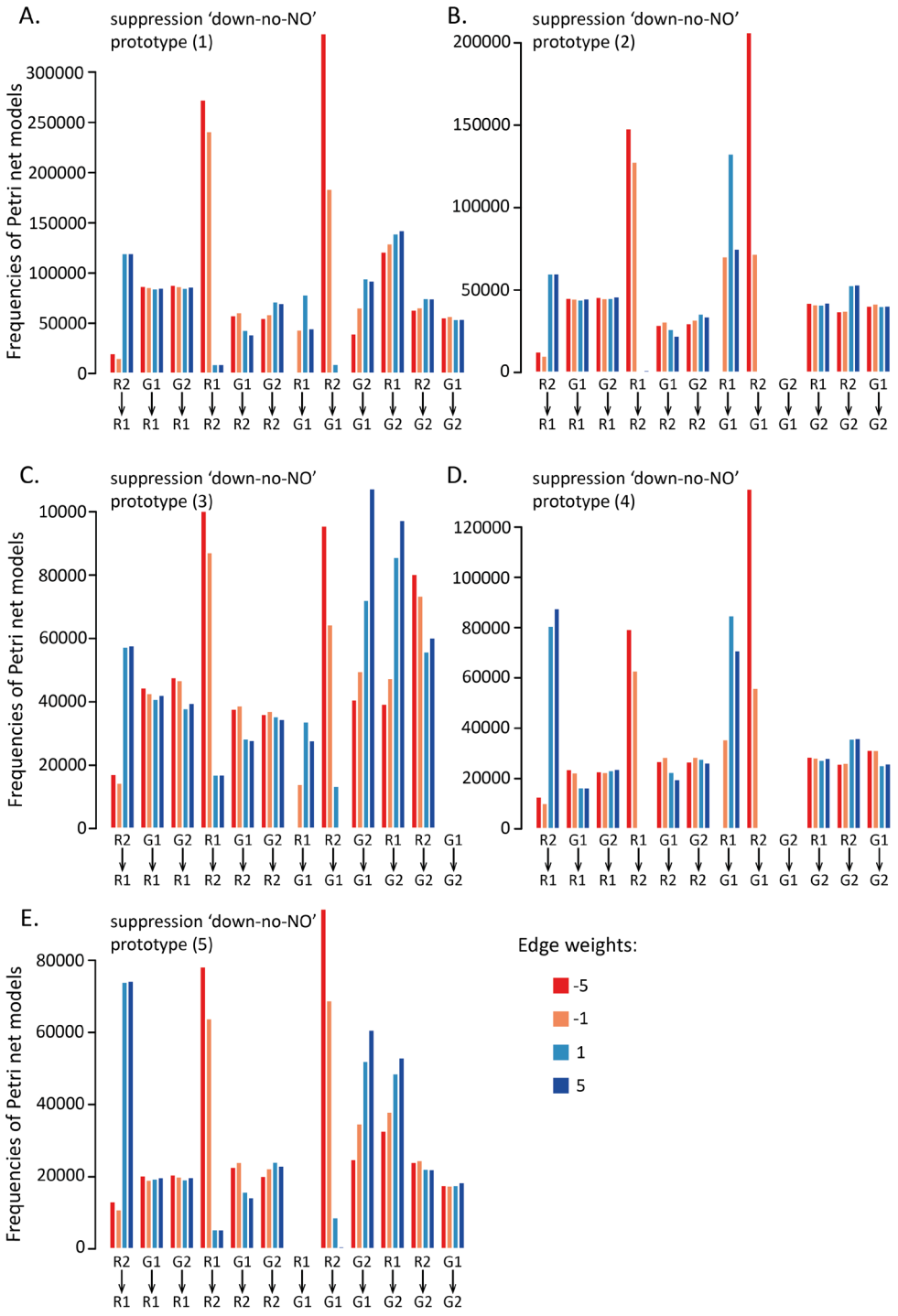
S5.6 Figure. Edge weight frequencies in Petri net models for the five most abundant prototypes (A-E) for inversion 'up-no-DOWN'. Red: -5, Orange: -1, light blue: 1, and dark blue: 5.



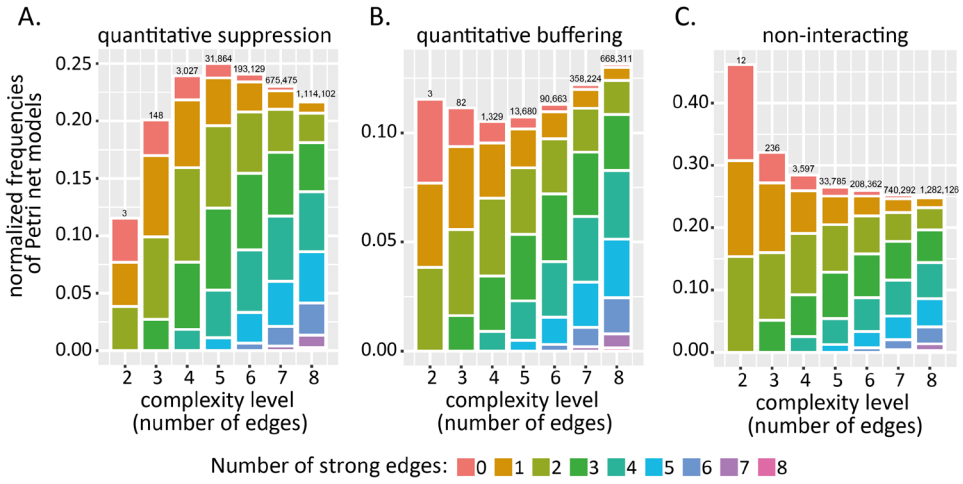
S5.7 Figure. Edge weight frequencies in Petri net models for the five most abundant prototypes (A-E) for inversion 'down-no-UP'. Red: -5, Orange: -1, light blue: 1, and dark blue: 5.



S5.8 Figure. Edge weight frequencies in Petri net models for the five most abundant prototypes (A-E) for suppression 'up-no-NO'. Red: -5, Orange: -1, light blue: 1, and dark blue: 5.



S5.9 Figure. Edge weight frequencies in Petri net models for the five most abundant prototypes (A-E) for suppression 'down-no-NO'. Red: -5, Orange: -1, light blue: 1, and dark blue: 5.



S5.10 Figure. Normalized frequencies of Petri net models and 'strong' edges per complexity level (number of edges) per genetic interaction pattern. The distributions are shown for A) quantitative suppression, B) quantitative buffering, and C) non-interaction. The number above each bar represents the actual number of Petri nets for that complexity. The colors represent the number of strong edges (weight +5/-5) in the Petri net.

CHAPTER 6

A framework for automatically generating executable pathway models specified in BioPAX

Bioinformatics. 2016, 32(12), i60-i69. doi: 10.1093/bioinformatics/btw250

Reza Haydarou, Annika Jacobsen, Nicola Bonzanni, K. Anton Feenstra, Sanne Abeln, Jaap Heringa

Abstract

Biological pathways play a key role in most cellular functions. To better understand these functions, diverse computational and cell biology researchers use biological pathway data for various analysis and modeling purposes. For specifying these biological pathways, a community of researchers has defined BioPAX and provided various tools for creating, validating and visualizing BioPAX models. However, a generic software framework for simulating BioPAX models is missing. Here, we attempt to fill this gap by introducing a generic simulation framework for BioPAX. The framework explicitly separates the execution model from the model structure as provided by BioPAX, with the advantage that the modelling process becomes more reproducible and intrinsically more modular; this ensures natural biological constraints are satisfied upon execution. The framework is based on the principles of discrete event systems and multi-agent systems, and is capable of automatically generating a hierarchical multi-agent system for a given BioPAX model.

To demonstrate the applicability of the framework, we simulated two types of biological network models: a gene regulatory network modeling the haematopoietic stem cell regulators and a signal transduction network modeling the Wnt/b-catenin signaling pathway. We observed that the results of the simulations performed using our framework were entirely consistent with the simulation results reported by the researchers who developed the original models in a proprietary language.

The framework, implemented in Java, is open source and its source code, documentation and tutorial are available at http://www.ibi.vu.nl/programs/BioASF.

6.1. Introduction

The National Human Genome Research Institute (NHGRI) has defined a biological pathway as 'a series of actions among molecules in a cell that leads to a certain product or a change in a cell' (http://www.genome.gov/27530687). Biological pathways play a key role in most cellular processes including metabolism, gene expression regulation and signal transduction. According to Pathguide (http://www.pathguide.org) (Bader et al., 2006), there are more than 500 resources, including databases, in which biological pathway data are stored in various data formats. To increase the uniformity of pathway data from different sources, make biological pathway data exchangeable, and increase the efficiency of computational pathway research, a community of researchers have defined BioPAX (Biological Pathway Exchange): a semantic-web based standard language to specify biological pathway models at the molecular and cellular level (Demir et al., 2010). A biological pathway specified in the BioPAX language is called a BioPAX model. Currently, the most comprehensive biological pathway databases, including Pathway Commons (Cerami et al., 2011), KEGG (Kanehisa and Goto, 2000), NCI/Nature PID (Schaefer et al., 2009), Reactome (Croft et al., 2014), WikiPathways (Kelder et al., 2012), and NetPath (Kandasamy et al., 2010), provide pathway descriptions in BioPAX format.

Since the introduction of BioPAX, the BioPAX community has provided various tools: Paxtools (Demir et al., 2013) for reading and writing BioPAX models, Validator (Rodchenkov et al., 2013) for validating a BioPAX model to see whether or not it complies with the BioPAX specification, Pattern Search for enabling the search of specific topological structures in a BioPAX model (Babur et al., 2014) and ChiBE (Babur et al., 2010) for visualizing BioPAX models in the standard Systems Biology Graphical Notation (SBGN) (Le Nove`re et al., 2009) format. Although simulation of biological pathway models makes it possible to achieve an adequate level of understanding of pathway models, a generic software framework for simulating biological pathway models specified in BioPAX has not been available to date. We address this omission by introducing BioASF (BioPAX-based Agent-oriented Simulation Framework).

BioASF is different in a number of ways from the existing simulation frameworks for biological pathways handling qualitative regulatory and signaling pathway models. These frameworks are mostly based on either Petri nets or process calculi. In a Petri net, a place can represent molecules such as genes, proteins or complexes, and a transition can represent their interactions. Firing of a transition leads to consuming substrates and creating products. The execution is constrained by the weight of an arc, which connects a place with a transition. An example of a Petri net based simulation framework is Cell Illustrator (Nagasaki et al., 2009). In process calculi, a communicating process represents molecules, a communication event represents an interaction and a state change represents a modification of the molecule. An example of a process calculi based simulation framework is Bio-PEPA

(Ciocchetta and Hillston, 2009). Note that because the focus of the simulation frameworks that are based on the Systems Biology Markup Language (SBML) (Hucka et al., 2003) is mainly the dynamic simulation of quantitative models, we do not mention these frameworks in this paper.

One important difference between Petri net and process calculibased frameworks and BioASF is the language used to specify pathway models. In a typical Petri net based simulation framework, a pathway model is specified in various proprietary languages in which molecules are assigned to places and molecular interactions are assigned to transitions. In a typical process calculi-based simulation framework, a pathway model is specified in a language containing terms such as actions and synchronizations. This language is primarily meant for specifying performance models of computer and communication systems. On the contrary, BioASF uses a standard language (BioPAX) for specifying pathway models.

Another important difference between the mentioned frameworks and BioASF is the type of network they build during simulation. Both Petri net and process calculi-based frameworks use intrinsically flat networks and the nature of the network nodes is very generic (places and transition nodes in a Petri net and processes and reaction nodes in process calculi). In BioASF the network nodes have a natural semantic meaning, are organized hierarchically, and the network exhibits a high degree of modularity. BioASF gets its modularity from BioPAX in which biological entities are defined in a hierarchical fashion, providing natural modularity for model developers.

Moreover, in BioASF, there is a clear distinction between the biological network model and simulation execution model. The biological network model in BioASF is represented as a network in which nodes are biological entities (such as proteins, genes, biochemical reactions, regulatory interactions), interconnected via well-defined properties and according to the constraints defined in the BioPAX language. Note that BioPAX is a specification language and it does not provide any execution semantics. In order to simulate BioPAX models, BioASF provides the simulation execution model (see Appendix). The simulation execution model is represented as a network in which nodes are agents controlling the execution of their corresponding biological interactions in the biological network model. The simulation execution model enforces three types of constraints during a model execution: *i*) constraints defined in BioPAX, *ii*) constraints defined by simulation rules and *iii*) constraints defined by analysis rules. Hence, BioASF explicitly requires the corresponding inputs. The explicit specification of these separate inputs enhances the reproducibility and consistency of the modeling procedure. For example, from a biological viewpoint, a protein cannot be translated from another protein. RNA is translated into proteins. BioPAX respects this constraint. The constraints provided by BioPAX disallows a template reaction interaction (an interaction that polymerizes its product based on a template, and is used for specifying transcriptions, translations and replications) to accept a protein or small molecule as its template. BioPAX requires the template to be either an RNA (for specifying translations) or DNA (for specifying transcriptions

and DNA replications). However, due to intrinsically very generic nature of nodes in Petri net (and also process calculi) frameworks, it is possible to define a node representing a protein and accidentally use this as a template for a transcription, translation, or a replication reaction.

Figure 6.1 depicts the different inputs and the output of BioASF. A BioASF user provides biological network data specified in BioPAX, a list of initial concentration values of the physical entities, simulation rules indicating how biological interactions should be performed, and analysis rules indicating how the simulation results over time are interpreted and analyzed. Based on these inputs, BioASF automatically generates a hierarchical multi-agent system where each agent in the system is a goal-directed and autonomous entity responsible for executing a BioPAX interaction and pathway. Throughout the simulation, the result of the execution of a BioPAX interaction influences the behavior of the agents and makes changes in the concentration of physical entities in the environment. During the simulation, it is possible to query BioASF about the current simulation data such as number and type of executed interactions and concentration level of physical entities. The following sections elaborate in more detail on the internal components, organization and various features of BioASF.

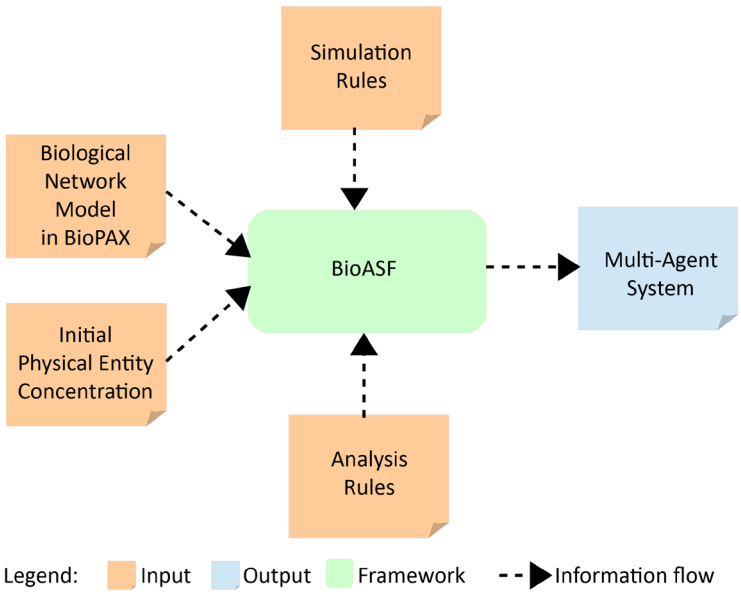


Figure 6.1. Inputs and output of BioASF. Inputs are a pathway description in BioPAX, initial concentration values of physical entities such as proteins, and rules governing the execution of interactions and the analysis of simulation results. The output is a multi-agent system which is automatically generated by BioASF and is responsible for executing BioPAX interactions.

6.2 Materials and methods

6.2.1 The BioPAX language

The BioPAX language is based on the Ontology Web Language (OWL) (Bechhofer, 2004), which is a description logic language intended to augment the current web with formalized knowledge in order to make information on the web machine-processable. Figure 6.2 depicts the BioPAX core concepts and their hierarchical relationships. Two main core concepts of the BioPAX language are pathway and interaction. A pathway is defined as a series of interactions demarcated by biologists in order to group interactions for certain biological reasons. A pathway can contain sub-pathways and interactions. An interaction represents a biological relationship among biological entities. A biological entity can be an interaction, pathway, gene, or physical entity (such as protein and DNA region). Interactions can transform one or more physical entities to others (conversion interactions) or control and regulate other interactions (control interactions). Interactions can be shared between different pathways (Demir et al., 2010).

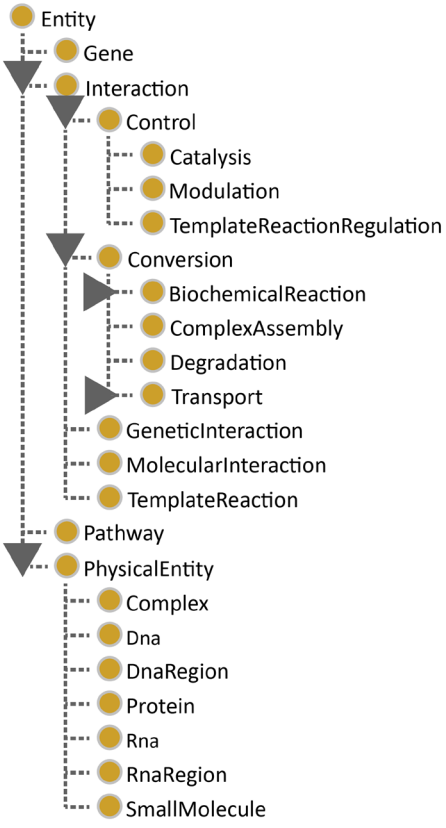


Figure 6.2. A snapshot from the Protégé OWL editor (http://protege.stanford.edu/overview/protege-owl.html) showing the core concepts of the BioPAX language and their hierarchical relationships.

There are four types of conversion interactions: 1) biochemical reaction representing a conversion in which molecules of physical entities undergo covalent modifications, 2) complex assembly representing a conversion in which either a set of physical entities aggregate to from a complex or a complex is disassembled into its constituents, 3) degradation representing a conversion in which a macro-molecule is degraded into its elementary units, and 4) transport representing a conversion in which a physical entity changes its sub-cellular location (Anwar, 2010).

BioPAX defines three types of control interactions: 1) catalysis representing an interaction in which a physical entity increases the rate of a conversion interaction, 2) modulation representing an interaction in which a physical entity modulates a catalysis interaction and 3) template reaction regulation representing an interaction in which a physical entity (such as a transcription factor) regulates the expression of a macro-molecule from a template macro-molecule (such as a DNA region) (Anwar, 2010). These BioPAX core concepts together with the BioPAX utility concepts and language properties form the BioPAX meta-model.

The basic idea behind BioASF is to associate an agent to each BioPAX pathway and BioPAX interaction in order to manage the execution of them. Note that execution of pathways and interactions change the concentration of physical entities in the environment. Additional agents are needed for managing the biological environment and analyzing changes in it. By representing execution managers as agents, we can base our framework on multi-agent technology. The challenge is to find a suitable communication pattern between different types of agents, as pathways are related to other pathways and interactions, and interactions are also related to other interactions and biological entities. We have formally specified all different types of agents, their inputs and outputs, and their communications in mathematical set notation (see Appendix).

6.2.2 Architecture of the framework

The framework is based on the principles of the discrete event systems (DES) theory (Cassandras et al., 2009) and the multi-agent systems (MAS) (Wooldridge, 2009) paradigm. Similar to a discrete event system, the various components of the framework cooperate with each other to perform a function by utilizing discrete events. Each component manages a number of discrete states, and the state transition depends on the occurrence of discrete events. Corresponding to a multi-agent system, each component of the framework is represented as an agent which is autonomous, exhibits goaldirected behavior, and operates according to its own rules. Figure 6.3 shows the event-based cooperation of different agents of the framework that realize a biological simulation.

A simulation is started by activation of the Analysis Agent that in turn activates the Environment Agent. This agent loads the initial concentrations of all relevant physical entities (such as proteins and small molecules) from a storage. After environment initialization, the Analysis Agent activates the master pathway agent (MPA). MPA is responsible for bootstrapping the simulation and controlling

all top-level pathway agents. Pathway agents in the framework are organized in a hierarchical fashion, with the MPA at the top of the hierarchy. There is also a hierarchical relationship between a pathway agent and interaction agents. Hence, both sub-pathway agents and interaction agents belonging to a pathway are considered the pathway's children. The MPA activates all top level pathway agents each of which in turn activates its immediate sub-pathway agents and its interaction agents. An activated pathway agent is now ready to listen to events coming from its children.

After the activation of an Interaction Agent is completed, the activated Interaction Agent subscribes itself to any changes in the concentration of particular physical entities published by the Environment Agent. An interaction agent receiving such an event checks its simulation rules to see whether the required condition for performing its associated BioPAX interaction can be met. For example, a biochemical reaction with two participants, each requiring a certain stoichiometry can only be performed if the concentrations of both participants are equal or higher than their stoichiometric value. Besides, if this biochemical reaction is controlled by a catalysis interaction then it must be checked to see whether the catalysis interaction has already occurred and whether its effector (i.e., its result) is available. As soon as the condition is met, the Interaction Agent performs its associated BioPAX interaction and sends its effector to the Environment Agent. At the same time, the Interaction

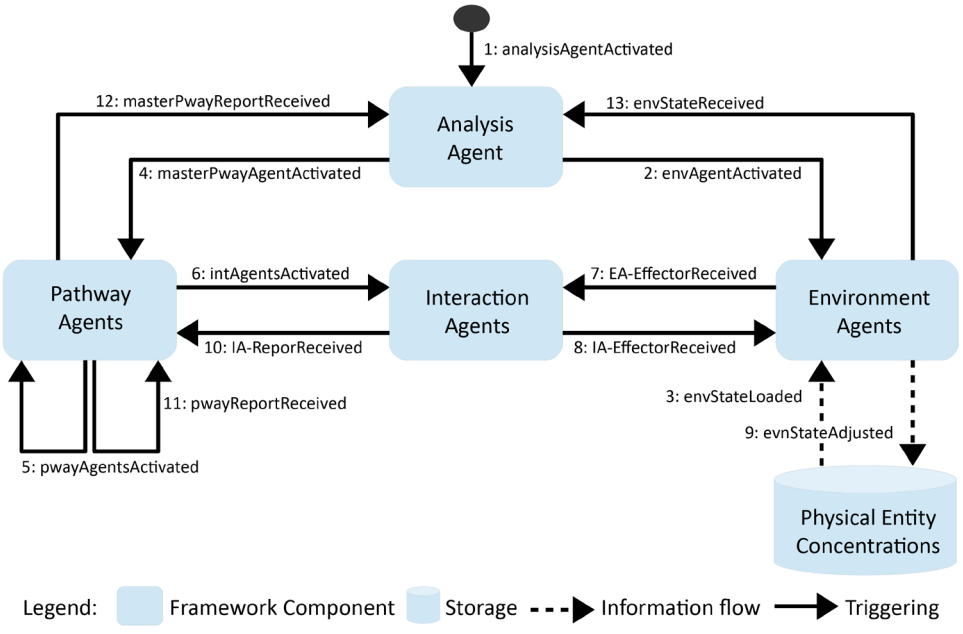


Figure 6.3. High level architecture of BioASF showing how different agents communicate with each other by exchanging events. Events are numbered chronologically.

Agent sends an event to its parent Pathway Agent. An example payload of such an event is a report indicating whether or not an interaction has occurred. The Pathway Agent integrates such reports received from its children and sends one integrated report to its parent. Finally, the MPA integrates all reports from the top level pathway agents and informs the Analysis Agent about the situation regarding the execution of all interactions. Consequently, Analysis Agent utilizes its analysis rules to store simulation results, examine environmental changes (e.g., changes in the concentration of physical entities), combine the current simulation results with the previous ones, and initiate a new cycle of simulation with new initial concentrations.

6.2.3 Artifacts of the framework

One of the important features of the framework is that it is able to generate pathway and interaction agents from a BioPAX model, and there is no need for users to write code for these agents. This feature is realized by utilizing various artifacts. Note that in software engineering, an artifact is a physical piece of information such as a data file, source code file, and library file that is used or produced during a software development process. We use this terminology to explain different software elements of the BioASF framework. depicted in Figure 6.4.

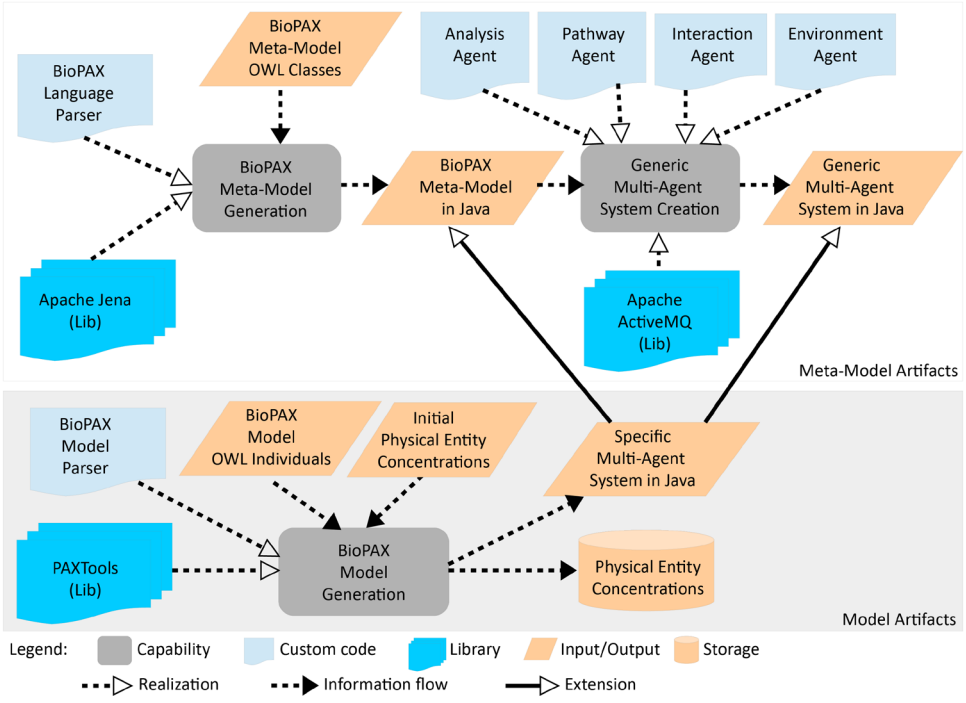


Figure 6.4. Meta-model layer and model layer artifacts of BioASF showing how the multi-agent system code in Java is generated for simulating the biological pathway specified in a given BioPAX model.

We make a distinction between meta-model artifacts and model artifacts. Meta-model artifacts are related to the BioPAX language constructs (OWL classes and properties) but model artifacts are related to particular pathways and interactions expressed in the BioPAX language (OWL individuals), which is also known as a BioPAX model.

Meta-model artifacts are obtained by two capabilities: BioPAX Meta-Model Generation and Generic Multi-Agent System Creation. The first capability takes as its input the BioPAX OWL classes and properties, parses the input by utilizing the Apache Jena library (Apache, 2009), and produces Java classes which are related to each other in the same way as the BioPAX OWL classes. The second capability uses the meta-level Java classes (which are the Java representation of the BioPAX meta-model) and the manually developed Java classes (which contain code for all different types of agents) to produce the generic multiagent system. This multi-agent system utilizes the messaging framework Apache ActiveMQ (Apache, 2005) for sending and receiving message events among agents. Also the Java classes of the multi-agent system are related to each other in the same fashion as the BioPAX OWL classes. The meta-model artifacts are created once in the life-cycle of the framework. They may require re-generation or adjustment if a new version of BioPAX is released.

Model artifacts are obtained by the capability BioPAX Model Generation which takes as its input the BioPAX OWL individuals, parses the input by utilizing the PAXTools library (Demir et al., 2013), and produces a specific multi-agent system which is an extension of the meta-level generic multi-agent system. Moreover, BioASF also generates a storage containing the concentration of all physical entities (such as proteins, complexes, small molecules) that are part of a BioPAX model.

6.2.4 Flexibility of the framework

The framework offers flexibility on the following four aspects: simulation rules, simulation modes, analysis mechanisms, and scalability. BioASF allows users to define their own simulation rules, choose their proper simulation mode, incorporate their own analysis mechanisms, and run simulation in a single or multiple Java Virtual Machine (JVM). The following four paragraphs briefly explain these aspects.

Obviously, it is impossible to have a one-size-fits-all set of rules governing all types of simulations. Each simulation requires its specific rules. For example, the authors in (Albert and Othmer, 2003) propose a Boolean model of regulatory interactions to get insights into the functioning of the segment polarity gene network in Drosophila. In their model, they represent mRNA or protein by a node and the interactions between them by a directed edge. The state of each node is determined by a Boolean function which is highly specific for each node and it does not follow a generic pattern. Another example is the network model of the cell cycle regulatory network of fission yeast (Davidich and Bornholdt, 2008). In the yeast cell cycle network, similar to the previous example, the nodes of the network are genes, and each node is assigned a binary value. Differently to the previous example, this value is determined by a generic rule for all nodes, except for a few nodes. However, this

generic rule is highly specific for the regulatory processes that control the cell cycle in fission yeast. Simulation rules are thus determined by the type of a biological network under study, the goal of an experiment, and the experimental data. Our framework provides well-defined programming interfaces that can be used to integrate specific simulation rules with the remaining parts of the framework.

BioASF supports three modes of simulations: synchronous (Garg et al., 2008), asynchronous (Garg et al., 2008), and maximally parallel (Burhard, 1983; Bonzanni et al., 2009b). The key difference between these modes is whether or not interaction agents within one execution step execute independently from each other. In synchronous mode, all interaction agents consume/produce physical entities simultaneously, and consumption/production of physical entities by one interaction agent has no influence on the other interaction agents. In asynchronous mode, interaction agents consume/produce physical entities in different times such that a physical entity produced by an interaction agent is immediately available for other interaction agents. Hence, they compete for consumption of the shared physical entities. In maximally parallel mode, interaction agents compete for consumption of physical entities but the production of one interaction agent is not immediately available for other interaction agents within the same execution step.

Each experiment requires its own approach for analyzing simulation results to get insight into the behavior of the model under study. For example, in an in silico experiment for cell fate determination of C. elegans during vulval induction, the simulation results are used as variables for specific scoring functions to assign to a cell the fate corresponding to the highest scoring function (Bonzanni et al., 2009b). In another in silico experiment for studying the behavior of the regulatory network of 11 haematopoietic stem/progenitor cell genes, the simulation results are used to find the steady state of the model by computing the terminal strongly connected components of the state space of the model (Bonzanni et al., 2013). These examples show that different analysis mechanisms may be used in each experiment. Our framework provides well-defined programming interfaces for integrating specific analysis mechanisms with the remaining parts of the framework.

Since a BioPAX model may contain multiple pathways each consisting of a large number of interactions, running simulation of such large models on one JVM (i.e., on one machine) can cause considerable problems. Therefore, BioASF provides two types of communication among agents: local communication using the Java observer and observable pattern (where a Java object (observable) maintains a list of its dependents (observers) and notifies them automatically of any state changes) and remote communication using the messaging system Apache ActiveMQ (where agents are distributed across a network and communicate with each other through queuing and publish/subscribe mechanisms).

6.3 Results

In order to validate BioASF, we have simulated two different biological network models in BioASF: a gene regulatory network modeling the haematopoietic stem cell (HSC) regulators (Bonzanni et al., 2013) and a signal transduction network modeling the Wnt/β-catenin signaling pathway (Jacobsen et al., 2016). Both original models have been developed in Petri net. We first made BioPAX models for both networks, used the BioASF's plugin to express the interaction rules required by the models, plugged the corresponding analysismodule in BioASF, prepared the initial concentration values for the physical entities occurring in the models, generated Java code for all agents, and finally ran the simulations. We observed that the results of the simulations performed using BioASF were similar to the results reported by the researchers in (Bonzanni et al., 2013) and (Jacobsen et al., 2016). In the following sections, we give a brief explanation of the simulations and show the results.

6.3.1 Gene regulatory simulation: model description

Haematopoiesis has long served as a model process for studying stem cells and represents the best characterized adult stem cell system. Transcriptional regulation is a key factor controlling haematopoiesis. In (Bonzanni et al., 2013), the authors build the first comprehensive regulatory network model based on the systematic curation of 11 fully validated regulatory elements linking together 11 transcription factors, all of which are active in early haematopoietic stem/progenitor cells. All 11 regulatory elements included in the model have been previously published and studied extensively using DNA/proteinbinding assays, as well as reporter gene assays of wild-type and mutant elements. The direction and value of each of the regulatory interactions are also known with certainty. Moreover, proteinprotein interactions curated from the literature were included. The resulting network was modeled as logical interactions encoding the activating and/or inhibitory links, including the specific combinations in which particular interactions occur.

6.3.2 Gene regulatory simulation: BioPAX representation

We considered the gene regulatory network controlling haematopoiesis as a BioPAX pathway and defined a BioPAX model. Figure 6.5 shows the graphical representation of the BioPAX model in SBGN format. In the model, we represented the gene product definitions for all genes used in the network as the BioPAX Protein, the transcription of the genes as the BioPAX TemplateReaction, and the regulatory role of the transcription factors in the transcription as the BioPAX TemplateReactionRegulation. For example, as shown in Figure 6.6, TmpReg-PU1-GATA1-GATA1 regulates (inhibits) the transcription of GATA1 (TmpReac-GATA1) through its controllers PU1 and GATA1.

6.3.3 Gene regulatory simulation: simulation rules

For the simulation of the haematopoiesis model, we extracted a number of simple rules from the description of the original in silico experiment (Bonzanni et al.,

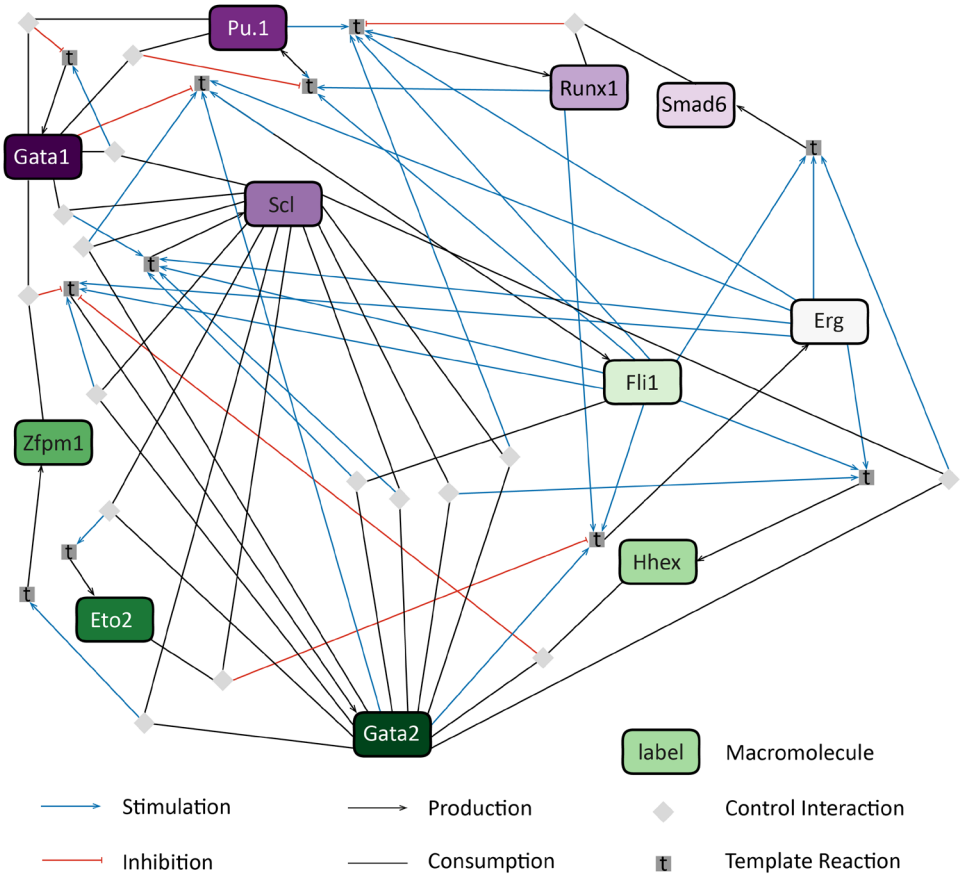


Figure 6.5. SBGN representation of the haematopoiesis gene regulatory network generated by ChiBE (Babur et al., 2010) based on our BioPAX model.

2009a, 2013), and plugged them in BioASF. We recognized three categories of rules: template reaction regulation rules, template reaction rules, and degradation rules. These rules determine whether or not a BioPAX interaction should be performed during the simulation. A template reaction regulation interaction can be performed if the concentration values of its inputs (i.e., all of its controllers) are higher than zero. After performing this interaction, the concentration value of its output is increased by one by the environment agent. There are two conditions which must be fulfilled before a template reaction interaction can be performed: (1) the template reaction regulation interaction regulating the template reaction interaction must already have occurred, and (2) the concentration value of the output (i.e., product) of the template reaction interaction in the environment must either be zero (in case, the control type is ACTIVATION) or one (in case, the control type is INHIBITION). After performing this interaction, the concentration value of its output is either increased or decreased by one by the environment agent, depending on the control

```
<bp:Protein rdf:ID="GATA1">
 <bp:displayName rdf:datatype = "&xsd;string">
     GATA1</bp:displayName>
</bp:Protein>
  ...
<bp:TemplateReaction rdf:ID="TmpReac-GATA1">
 <bp:product rdf:resource="#GATA1"/>
</bp:TemplateReaction>
  ...
<bp:TemplateReactionRegulation
           rdf:ID="TmpReg-PU1-GATA1-GATA1">
 <bp:controlType rdf:datatype="&xsd;string">
           INHIBITION</bp:controlType>
 <bp:controller rdf:resource="#PU1"/>
 <bp:controller rdf:resource="#GATA1"/>
 <bp:controlled rdf:resource="#TmpReac-GATA1"/>
</bp:TemplateReactionRegulation>
  ...
<bp:Pathway rdf:about="Haematopisis">
  <bp:pathwayComponent rdf:resource="#TmpReac-GATA1" />
  <bp:pathwayComponent rdf:resource=
           "#TmpReg-PU1-GATA1-GATA1" />
  ...
```

Figure 6.6. Part of the BioPAX model for the haematopoiesis gene regulatory network.

type of its regulator. According to the degradation rules, the output of a template reaction interaction is degraded if the concentration value of one of its inputs in the environment is either zero (in case, the control type is ACTIVATION) or one (in case, the control type is INHIBITION).

As the firing of an interaction in the haematopoiesis Petri net model has no impact on the other interactions, within one execution step, this model was simulated in BioASF in the synchronous mode.

6.3.4 Gene regulatory simulation: model analysis

For the analysis of the simulation results of the haematopoiesis model, we created an analysis module in Java and plugged it in BioASF. This module initially constructs the state space graph of 2048 nodes each containing a bit vector of size 11, initiates the concentration value of all gene products (0 or 1) at the beginning of each simulation step for each state, utilizes BioASF to execute one simulation step, and updates the graph based on the simulation results. In line with the original in silico experiment (Bonzanni et al., 2009a, 2013), at the end of the simulation, after the graph was completely built, the terminal strongly connected components (TSCC) of the state space are computed in order to identify the attractors which represent the biological steady states.

Table 6.1. The result of the analysis of the haematopoiesis simulation. Each row shows a TSCC (i.e. stable state) in which genes (the columns 2–11) have been switched on (1) or off (0)

Comp	Erg	Eto2	Fli1	GATA1	Gata2	Hhex	Pu.1	Runx1	Scl	Smad6	Zfpm1
TSCC-1	0	0	0	0	0	0	0	0	0	0	0
TSCC-2	0	0	0	1	0	0	0	0	1	0	0
TSCC-3	0	0	1	0	0	1	1	0	1	1	0
TSCC-3	1	0	1	0	0	1	1	0	1	1	0
TSCC-3	0	1	1	0	0	1	1	0	1	1	0
TSCC-3	1	1	1	0	0	1	1	0	1	1	0
TSCC-3	0	0	1	0	1	1	1	0	1	1	0
TSCC-3	1	0	1	0	1	1	1	0	1	1	0
TSCC-3	0	1	1	0	1	1	1	0	1	1	0
TSCC-3	1	1	1	0	1	1	1	0	1	1	0
TSCC-3	0	0	1	0	0	1	1	1	1	1	0
TSCC-3	1	0	1	0	0	1	1	1	1	1	0
TSCC-3	0	1	1	0	0	1	1	1	1	1	0
TSCC-3	1	1	1	0	0	1	1	1	1	1	0
TSCC-3	0	0	1	0	1	1	1	1	1	1	0
TSCC-3	1	0	1	0	1	1	1	1	1	1	0
TSCC-3	0	1	1	0	1	1	1	1	1	1	0
TSCC-3	1	1	1	0	1	1	1	1	1	1	0
TSCC-3	0	0	1	0	0	1	1	0	1	1	1
TSCC-3	1	0	1	0	0	1	1	0	1	1	1
TSCC-3	0	1	1	0	0	1	1	0	1	1	1
TSCC-3	1	1	1	0	0	1	1	0	1	1	1
TSCC-3	0	0	1	0	1	1	1	0	1	1	1
TSCC-3	1	0	1	0	1	1	1	0	1	1	1
TSCC-3	0	1	1	0	1	1	1	0	1	1	1
TSCC-3	1	1	1	0	1	1	1	0	1	1	1
TSCC-3	0	0	1	0	0	1	1	1	1	1	1
TSCC-3	1	0	1	0	0	1	1	1	1	1	1
TSCC-3	0	1	1	0	0	1	1	1	1	1	1
TSCC-3	1	1	1	0	0	1	1	1	1	1	1
TSCC-3	0	0	1	0	1	1	1	1	1	1	1
TSCC-3	1	0	1	0	1	1	1	1	1	1	1
TSCC-3	0	1	1	0	1	1	1	1	1	1	1
TSCC-3	1	1	1	0	1	1	1	1	1	1	1

As reported by the original in silico experiment, we found three TSCCs: the TSCC-1 containing one state that corresponds to a nonhaematopoietic cell, the TSCC-2 containing one state that closely resembles a mature erythrocyte, and the TSCC-3 containing 32 states that match the expected pattern for haematopoietic stem cells, as shown in Table 6.1.

6.3.5 Signal transduction simulation: model description

Wnt/β-catenin signaling is highly conserved in all animals and is important for embryonic development and stem cell maintenance in adults. WNT ligand stimulation leads to inhibition of an important destruction complex that degrades CTNNB1 (β-catenin). This leads to CTNNB1 accumulation and translocation to the nucleus, where it binds TCF/LEF and activates transcription ofWNT target

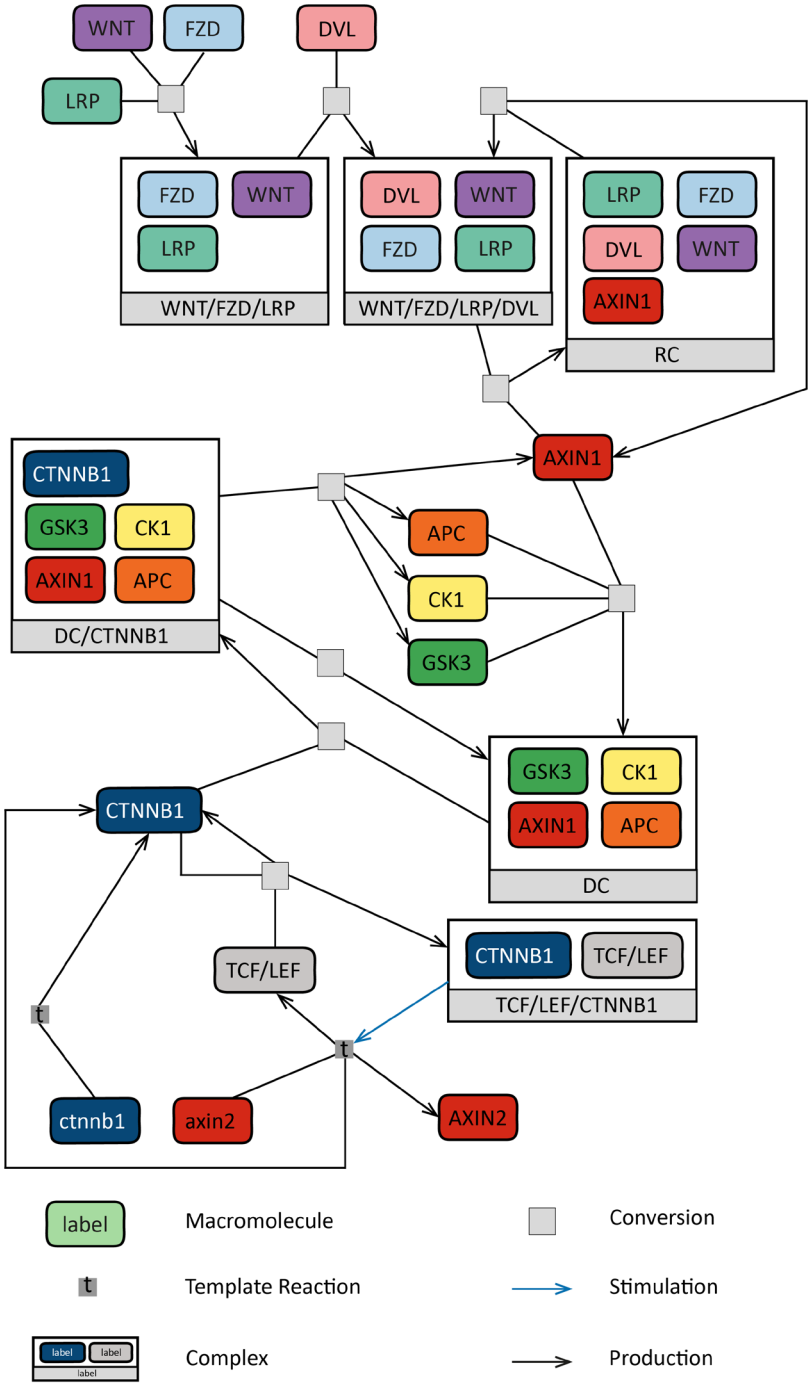


Figure 6.7. SBGN representation of the WNT/ β-catenin signaling pathway generated by ChiBE (Babur et al., 2010) based on our BioPAX model. The highlighted part of the graph corresponds to the part of the BioPAX model depicted in Figure 6.5.

genes. In (Jacobsen et al., 2016), the authors build a Wnt/β-catenin signaling model based on interactions, well established in literature. Signaling is initiated by the binding of external WNT ligand to the receptor, FZD, and the co-receptor, LRP. Dishevelled (DVL) and AXIN bind the intracellular tail of FZD. AXIN, APC, CK1 and GSK3 form the destruction complex. CTNNB1 is constitutively produced and either binds the destruction complex for degradation or translocates to the nucleus, where it activates transcription of WNT target genes.

6.3.6 Signal transduction simulation: BioPAX representation

Figure 6.7 shows the graphical representation of the BioPAX model for the Wnt/β-catenin signaling pathway in SBGN format. In the BioPAX model , we used a number of BioPAX classes including DnaRegion (for specifying the genes such as

```
<bp:ComplexAssembly rdf:ID="CmpAsm-TCF-LEF-CTNNB1">
 <bp:left rdf:resource="#CTNNB1" />
 <bp:left rdf:resource="#TCF-LEF" />
 <bp:participantStoichiometry rdf:resource=
      "#Stoi-CmpAsm-Left-CTNNB1" />
 <bp:right rdf:resource="#TCF-LEF-CTNNB1" />
 <bp:right rdf:resource="#CTNNB1" />
 <bp:participantStoichiometry rdf:resource=
      "#Stoi-CmpAsm-Right-CTNNB1" />
</bp:ComplexAssembly>
<bp:Stoichiometry rdf:ID="Stoi-CmpAsm-Left-CTNNB1">
 <bp:physicalEntity rdf:resource="#CTNNB1" />
 <bp:stoichiometricCoefficient rdf:datatype =
      "&xsd;float">3.0</bp:stoichiometricCoefficient>
</bp:Stoichiometry>
<bp:Stoichiometry rdf:ID="Stoi-CmpAsm-Right-CTNNB1">
 <bp:physicalEntity rdf:resource="#CTNNB1" />
 <bp:stoichiometricCoefficient rdf:datatype =
      "&xsd;float">2.0</bp:stoichiometricCoefficient>
</bp:Stoichiometry>
<bp:TemplateReactionRegulation rdf:ID=
      "TmpReg-TCF-LEF-CTNNB1-TARGET">
 <bp:controlType rdf:datatype="&xsd;string">
      ACTIVATION</bp:controlType>
 <bp:controller rdf:resource="#TCF-LEF-CTNNB1"/>
 <bp:controlled rdf:resource="#TmpReac-TARGET"/>
</bp:TemplateReactionRegulation>
  ...
<bp:Pathway rdf:about="Wnt-Experiment">
  ...
  <bp:pathwayComponent rdf:resource=
      "#CmpAsm-TCF-LEF-CTNNB1" />
  <bp:pathwayComponent rdf:resource=
      "#TmpReg-TCF-LEF-CTNNB1-TARGET" />
  ...
```

Figure 6.8. Part of the BioPAX model for the WNT/β-catenin signaling pathway.

ctnnb1), ComplexAssembly (for the formation of a complex from a number of proteins such as the formation of the destruction complex from AXIN, APC, CK1 and GSK3), and Degradation (for specifying the degradation of CTNNB1).

Figure 6.8 shows a part of the BioPAX model in which the complex assembly interaction CmpAsm-TCF-LEF-CTNNB1 uses CTNNB1 and TCF-LEF to create the TCF-LEF-CTNNB1 complex. This complex plays the role of transcription factor in the template regulation interaction TmpReg-TCF-LEF-CTNNB1-TARGET, which regulates (activates) the transcription of the WNT target genes. In order to slow down the rate of the complex assembly interaction and the transport of CTNNB1 from the nuclear to the cytoplasm, we used the BioPAX Stoichiometry class.

6.3.7 Signal transduction simulation: simulation rules

Similar to the haematopoiesis simulation, for the simulation of the WNT/β-catenin model, we extracted a number of rules from the description of the original in silico experiment (Jacobsen et al., 2016), and plugged them in BioASF. We recognized three categories of rules: template reaction regulation rules, template reaction rules, and conversion rules. The conversion rules are applicable for all complex assembly and degradation interactions. We also defined a generic rule which determines whether or not an interaction is allowed to be executed. According to this rule, an interaction might be performed if the stoichiometries of all participants of an interaction are equal or higher than the available concentration values of the participants. In keeping with the Petri net formalism, the stoichiometry of a participant is assumed to be one if it has not been specified.

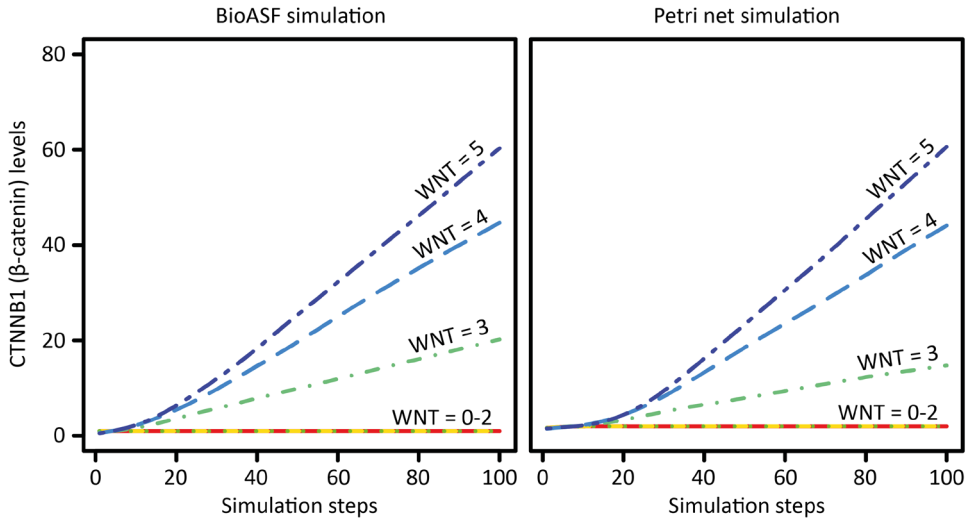


Figure 6.9. The CTNNB1 (β-catenin) levels for initial WNT levels from 0 to 5 simulated 50 times over 100 steps using BioASF and Petri-net (Jacobsen et al., 2016), respectively

Another generic rule is that the concentration value of the output of an interaction is increased by one by the environment agent, after the interaction was performed. The changes in the concentration value of the input of an interaction depend on its type. The concentration values of the inputs of the template reaction regulation and template reaction interactions are decreased by one, and the ones of the conversion interactions are decreased by the stoichiometry of their participants. As the firing of an interaction in the WNT/β-catenin model leads to the consumption of its input physical entities and they are not any more available for other interactions, within the same execution step, this model was simulated in BioASF in the maximally parallel mode.

6.3.8 Signal transduction simulation: model analysis

Similar to the gene regulatory simulation, for the analysis of the simulation results of the WNT/β-catenin model, we created an analysis module in Java and plugged it in BioASF. This module initially populates the pre-defined concentration values of all proteins, complexes, and genes, creates a file for writing the results of the simulations, utilizes BioASF to execute one simulation step, and writes the CTNNB1 levels in the file. In line with the original in silico experiment (Jacobsen et al., 2016), the model is simulated with initial WNT levels ranging from 0 to 5 over 100 steps. The simulation for each initial WNT level is repeated 50 times and the CTNNB1 levels for each step is stored in a file. Finally, an R module uses this file to calculate, for each initial WNT level, the mean CTNNB1 level for each step over the 50 simulations and plots these.

The BioASF simulations closely resemble the increase in the CTNNB1 levels from the original Petri net simulations, as shown in Figure 6.9. Similar to the original simulations, we observe four different CTNNB1 levels in response to the initial WNT levels (0 to 5): a horizontal line in response to the WNT levels of 0 to 2, and three oblique lines with three slopes in response to the initial WNT levels of 3 to 5.

6.4 Discussion

We introduced here a simulation software framework, called BioASF, for simulating biological pathway models specified in BioPAX. We demonstrated the applicability of BioASF by simulating models of haematopoiesis and Wnt/β-catenin signaling, which reproduced the results reported by the researchers who developed the original models. These two models had already been developed in our group and therefore the exact data of the original in silico experiments (initial concentration, simulation and analysis rules) were available. The availability of this data is required in order to compare the simulation results of BioASF with the results of the original ones. Although obtaining such input data to compare modelling frameworks may be difficult, particularly when originating in other groups, we feel that the clarity of BioASF and its explicit input requirements safeguard sound simulation of pathway models developed by other groups.

The main advantages of BioASF are: 1) ability to directly simulate pathway models specified in the standard language for pathway descriptions (i.e., BioPAX), 2) support for hierarchical (nested) pathway simulations, 3) inherent scalability because of using agent technology, 4) extensibility by providing pluggable architecture, 5) provision of flexibility to the biological model developers by enforcing the separation of concerns principles (i.e., a clear distinction is made among biological model, execution model, simulation rules, and simulation result analysis), and 6) formal specification of the execution model in mathematical set notation. The current BioASF limitations are: 1) disability to simulate differential equations, 2) required rules can only be specified in Java, and 3) no user interface for starting/stopping/debugging simulations. This makes the framework well suited for developers, but not yet easily accessible for biologists. Nevertheless, the website includes an extended tutorial to help non-expert users getting started.

As mentioned in Section 2.4, each simulation requires its specific rules on pathway and interaction levels, governing execution of a pathway or interaction. In BioASF, these rules currently are specified in Java, which negatively influences the framework's usability and deployability. The preferred way would be to express simulation rules in a rule language such as Semantic Web Rule Language (SWRL) (Horrocks et al., 2004) and relate these rules to the biological interactions in BioPAX models. This is similar to the approach taken in the Semantic Markup for Web Services (OWL-S) (Martin et al., 2004) where the pre-conditions of an OWL-S service can be expressed in a rule language. In this way, both biological models and simulation rules can be provided by the same users (e.g., biomedical researchers).

The BioASF framework can be used in different scenarios. For example, the framework can be used for model calibration. In this scenario, a BioPAX model can be visualized in SBGN and shown to users. Users can define break-points and watchpoints on pathways and interactions, similar to the debugging of application codes. When a break-point is hit, users can then inspect the concentration value of the related physical entities and they can change these values during the simulation in order to influence the simulation results and consequently adjust their model. Another possibility would be to use BioASF as a tool for measuring the fitness of individuals in the population in a genetic programming experiment. The purpose of such an experiment would be to automatically generate biological pathways based on a combination and integration of gene expression data, protein-to-protein interaction data, existing pathway data, gene ontology data, and triplets representing other biological data. Our future work will be focused on these items.

References

Albert R, Othmer HG. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol*. 2003, 223, 1-18.

Anwar N, et al. BioPAX Biological Pathways Exchange Language Level 3, Release Version 1 Documentation. *BioPAX Workgroup*. 2010.

Apache Software Foundation. Apache ActiveMQ: a powerful open source messaging and Integration Patterns server. 2005.

Apache Software Foundation. A free and open source Java framework for building Semantic Web and Linked Data applications. 2009.

Babur Ö, Dogrusoz U, Demir E, Sander C. ChiBE: interactive visualization and manipulation of BioPAX pathway models. *Bioinformatics*. 2010, 26, 429-431.

Babur Ö, Aksoy BA, Rodchenkov I, Sümer SO, Sander C, Demir E. Pattern search in BioPAX models. *Bioinformatics*. 2014, 30, 139-140.

Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. *Nucleic Acids Res*. 2006, 34, D504-D506.

Bechhofer S, van Harmelen F, Hendler J, Horrocks J, McGuinness DL, Patel-Schneider PF, et al. OWL Web Ontology Language Reference. World Wide Web Consortium. 2004.

Bonzanni N, Feenstra KA, Fokkink W, Krepska E. What Can Formal Methods Bring to Systems Biology? FM 2009: Formal Methods. Springer, Berlin, Heidelberg; 2009. pp. 16-22.

Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. *Bioinformatics*. 2009b, 25, 2049-56.

Bonzanni N, Garg A, Feenstra KA, Schütte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013, 29, i80-i8.

Burhard HD. On priorities of parallelism: Petri nets under the maximum firing strategy. In: Logics of Programs and Their Applications, vol. 148. Springer, Berlin/Heidelberg, pp. 15, 18, 32. 1983.

Cassandras CG, Lafortune S. Introduction to Discrete Event Systems. Springer. 2009, 2, 2-35.

Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway Commons: a web resource for biological pathway data. *Nucleic Acids Res*. 2011, 39, D685-D90.

Ciocchetta F, Hillston J. Bio-PEPA: a framework for the modelling and analysis of biological systems. *Theor. Comput. Sci*. 2009, 410, 3065-84.

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014, 42, D472-D7.

Davidich MI, Bornholdt S. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One*. 2008, 3, e1672.

Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, et al. The BioPAX community standard for pathway data sharing. *Nat. Biotechnol*. 2010, 28, 935-42.

Demir E, Babur Ö, Rodchenkov I, Aksoy BA, Fukuda KI, Gross B, et al. Using biological pathway data with paxtools. *PLoS Comput Biol*. 2013, 9, e1003194.

Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G. Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics*. 2008, 24, 1917-25.

Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M. SWRL: A Semantic Web Rule Language Combining OWL And RuleML. World Wide Web Consortium. 2004.

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003, 19, 524-31.

Jacobsen A, Heijmans N, Verkaar F, Smit MJ, Heringa J, van Amerongen R, et al. Construction and Experimental Validation of a Petri net model of Wnt/β-catenin signaling. *PLoS One*. 2016, 11, e0155743. **Chapter 2 of this thesis.**

Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, et al. NetPath: a public resource of curated signal transduction pathways. *Genome Biol*. 2010, 11, R3.

Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000, 28, 27-30.

Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: building research communities on biological pathways. *Nucleic Acids Res*. 2012, 40, D1301-D7.

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The systems biology graphical notation. *Nat. Biotechnol*. 2009, 27, 735-41.

Martin D, Burstein M, Hobbs J, Lassila O, McDermott D, McIlraith S, et al. OWL-S: Semantic Markup for Web Services. World Wide Web Consortium. 2004

Nagasaki M, Saito A, Doi A, Matsuno H, Miyano S. Foundations of systems biology. Using Cell Illustrator and Pathway Databases. Springer, London, p. 16. 2009

Rodchenkov I, Demir E, Sander C, Bader GD. The BioPAX Validator. *Bioinformatics*. 2013, 29, 2659-60.

Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, et al. PID: the pathway interaction database. *Nucleic Acids Res*. 2009, 37, D674-D9.

Wooldridge M. An Introduction to MultiAgent Systems, vol. 2. Wiley, Glasgow, pp. 21-38. 2009.

Appendix: Formal specification of BioASF

Inspired by the description logic of BioPAX, we define our multiagent system model (MASM) as a tuple. Note that here, we follow the mathematical convention to use capital letters to denote a set (e.g. I), and lower-case letters to denote one single element (e.g. i). Besides, we denote all agents with the superscript a (e.g. I^a or i^a). The tuple is defined as follows:

$$MASM =< mp^a, P^a, I^a, e^a, a^a, B, S, E, T, R >$$

where mp^a is the *master pathway agent* which is responsible for bootstrapping the simulation and controlling all top-level *pathway agents*. P^a is a set of pathway agents, each executing its associated pathway and controlling its sub-pathway agents and *interaction agents*. I^a is a set of interaction agents executing interactions. e^a is the *environment agent* maintaining and monitoring the status of physical entities, and a^a is the *analysis agent* responsible for analyzing the new situation that arises as a result of execution of pathways and interactions. A set of *biological entities* is represented by B. S is a set of *sensors* which trigger an agent to start its execution, and E is a set of *effectors* which is the result of the execution of an agent and which can influence the execution of other agents. T is a set of *transition rules* dictating the execution pattern of an agent. How and whether or not an agent should be executed are determined by these rules. This element of the model allows for model flexibility and extensibility, as model users can provide their own rules. Finally, R is a set of *relations* among pathways, interactions and physical entities.

In line with the BioPAX description logic, the B set can be partitioned as follows:

$$B = G \cup I \cup P \cup H$$

where G is a set of *genes*, I is a set of *interactions*, P is a set of *pathways*, H is a set of *physical entities*. Accordingly, the I set can be partitioned as follows:

$$I = CI \cup VI \cup GI \cup MI \cup TI$$

where CI is a set of *control interactions*, VI is a set of *conversion interactions*, GI is a set of *genetic interactions*, MI is a set of *molecular interactions*, and TI is a set of *template reactions*. Furthermore, CI is partitioned as follows:

$$CI = CCI \cup TCI \cup MCI$$

where CCI is a set of *catalysis interactions*, TCI is a set of *template reaction regulation interactions*, and MCI is a set of *modulation interactions*. The set of conversion interactions is also partitioned:

$$VI = AVI \cup BVI \cup DVI \cup TVI$$

where *AVI* is a set of *complex assembly interactions*, *BVI* is a set of *biochemical reaction interactions*, *DVI* is a set of *degradation interactions*, and *TVI* is a set of *transport interactions*.

Corresponding to each BioPAX interaction type, the model of each agent is formally specified. Each agent is considered as a simple entity having knowledge only about its own managed entity (a BioPAX interaction or pathway) and about its limited relationships with other related agents. Note that there are no agents associated with genes and physical entities in our model. We have specified the formal models of all agent types. Here, for the sake of illustration, we describe the model of pathway agent and catalysis agent.

Each member p^a of the P^a element of our multi-agent system model (*MASM*) is also defined as a tuple in which it manages its associated BioPAX pathway p_p, is triggered by its sensors S_p, influences other agents by its execution results E_p, and is governed by its rules T_p:

$$p^a = < p_p, I^a_p, P^a_p, S_p, E_p, T_p >$$

where $p_p \in P$, $I^a_p \subset I^a$, $P^a_p \subset P^a$, $S_p \subset S$, $E_p \subset E$, and $T_p \subset T$. From the viewpoint of a pathway agent, interactions belonging to its managed pathway are called *internal interactions* and the ones belonging to some other pathway are called *external interactions*. A pathway agent can be triggered by agents managing its internal interactions, sub-pathways, or external control interactions, all together forming the sensor set of the pathway agent (with $i \in I$ and $ci \in CI$):

$$S_p = \left\{ i^a \in I^a_p \mid managed(i^a, i) \wedge component(p_p, i) \right\} \cup$$

$$\left\{ P^a_s \in P^a_p \mid managed(P^a_s, P_s) \wedge component(p_p, P_s) \right\} \cup$$

$$\left\{ ci^a \in CI^a \mid managed(ci^a, ci) \wedge component(ci, p_p) \right\}$$

Similarly, a pathway agent can influence agents managing external interactions or external control interactions by playing the role of *participant* or *controller* respectively:

$$E_p = \left\{ i^a \in I^a_p \mid managed(i^a, i) \wedge participant(i, p_p) \right\} \cup$$

$$\left\{ ci^a \in CI^a S \mid managed(ci^a, ci) \wedge controller(ci, p_p) \right\}$$

Each catalysis agent (cci^a) is defined as a tuple in which it manages its associated BioPAX catalysis interaction (cci_c), is triggered by its sensors (S_c), influences other agents by its execution results (E_c), and is governed by its rules (T_c):

$$cci^a = < cci_c,\ S_c,\ E_c,\ T_c >$$

where $cci_c \in CCI$, $S_c \subset S$, $E_c \subset E$ and $T_c \subset T$. A catalysis agent can be triggered by the environment agent maintaining concentration of its controllers, and by modulation agents managing its modulation interactions (with $h \in H$ and $m \in MCI$):

$$S_c = \left\{ e^a \mid managed(e^a, h) \wedge controller(cci_c, h) \right\} \cup$$

$$\left\{ e^a \mid managed(e^a, h) \wedge cofactor(cci_c, h) \right\} \cup$$

$$\left\{ mci^a \in MCI^a \mid managed(mci^a, mci) \wedge controlled(mci, cci_c) \right\}$$

A catalysis agent can influence agents managing its conversion interactions (with $vi \in VI$):

$$E_c = \left\{ vi^a \in VI^a \mid managed(vi^a, vi) \wedge controlled(cci_c, vi) \right\}$$

CHAPTER 7

Discussion

In this thesis we have demonstrated how network models can be used to reason about molecular mechanisms of a cell. In Chapter 2 we presented a Petri net model of Wnt/β-catenin signaling that we used to investigate the sequence of events in physiological and aberrant signaling occurring in the early development of colorectal cancer. In Chapter 3 we proposed deregulated signaling mechanisms during further colorectal cancer progression. In particular, we described deregulated genes connecting the cell cycle regulator, Aurora kinase A (AURKA), and the Wnt/β-catenin and Ras-MAPK signaling pathways by applying a network-aware analysis to gene expression data. In Chapter 4 we compared genetic interaction patterns for gene specific transcription factors and genes related to signaling in yeast. We then exhaustively modeled these patterns using Petri nets deriving plausible mechanisms underpinning their regulation. The framework used in this approach is described in detail in Chapter 5. In Chapter 6 we described a modeling framework for simulating predefined biological models and specify the requirements for generating an executable model.

7.1 Molecular network models

Molecular data can be described in abstract representations, such as networks, where the nodes are molecules and the edges are molecular interactions. Molecular networks, such as gene regulatory, protein-protein interaction, signaling, metabolic, and genetic interaction networks, may provide essential insight into their components, pattern of connectivity and functions, and help us to reason about the role of individual components in health and disease.

To study and analyze molecular networks, many network-aware approaches have been developed that integrate independent molecular data sources, such as gene expression data, to these networks (Dittrich et al., 2008; Torkamani and Schork, 2009; Vandin et al., 2011; Ciriello et al., 2012). Here, sets of genes connected in the network (gene modules) enriched for a specific property in the integrated data source, e.g. significance in deregulation for gene expression data, are identified. These gene modules can be considered 'functionally' deregulated and to have an important role in the regulatory mechanism of the phenotype being study. This is in particular relevant for the study of cancer, where the mutations are found in a few specific signaling pathways (Vogelstein and Kinzler, 2004) and therefore would be expected to cluster together in modules.

To investigate how molecules interact over time different modeling techniques can be used that simulate dynamical behavior, such as Boolean networks (Kauffman, 1969), Petri nets (Petri, 1962), and ordinary differential equations (ODEs) (Goodwin,

1963) (Karlebach and Shamir, 2008). We have previously used Petri nets to model biological networks (Bonzanni et al., 2009; Bonzanni et al., 2009; Bonzanni et al., 2013), and in this thesis we have applied Petri nets in two separate studies, to model Wnt/β-catenin signaling in Chapter 2 and to exhaustively model genetic interaction patterns in Chapters 4 and 5.

Network models are often biased by available knowledge due to trending research interest. Moreover, the perception of knowledge also differs between individual researchers within and between different biological fields. This can, however, be an advantage when network models are generated through interdisciplinary collaborations, which in our case has resulted in insightful discussions both in formulating the research questions, and in the transition from molecular data to a formal description (Chapter 2, 3, and 4). Here, by generating and studying molecular network models we reasoned about underlying mechanisms of deregulated Wnt/β-catenin signaling (Chapter 2), provided alternative putative mechanisms for the interconnection between AURKA signaling and Wnt and Ras-MAPK signaling (Chapter 3), and exhaustively modeled genetic interactions to derive plausible molecular mechanisms (Chapter 4).

7.2 Petri net modeling: Wnt/β-catenin signaling

Wnt/β-catenin signaling is highly conserved and plays a crucial role in the regulation of cell proliferation and differentiation during development, as well as during tissue maintenance in adults (Cadigan and Peifer, 2009; Clevers and Nusse, 2012). Deregulation of Wnt/β-catenin signaling, caused by alterations of key pathway proteins is involved in the initial step of colorectal cancer formation where normal epithelium is transformed to adenomatous tissue. Wnt/β-catenin signaling is initiated by the extracellular stimulation of Wnt ligands, followed by a cascade of events involving multiple components and concludes with β-catenin accumulating in the cytoplasm. The current knowledge describing the molecular mechanisms of Wnt/β-catenin signaling is the result of more than 30 years of research (Nusse and Varmus, 2012). The Petri net model of Wnt/β-catenin signaling described in Chapter 2 is based on state-of-the-art knowledge in the field and represents a formal description of how we currently understand the pathway, which we could validate by comparing the model predictions with experimental results.

The model allowed us to reason about the molecular mechanisms leading to β-catenin stabilization in physiological and aberrant Wnt/β-catenin signaling, where two important mechanisms emerged. First, the negative feedback regulator, AXIN2, is only able to attenuate Wnt-pathway activity during physiological Wnt/β-catenin signaling, where AXIN1 is the limiting factor. Thus AXIN2 has no effect on alternative perturbations leading to deregulated Wnt/β-catenin signaling. Second, physiological Wnt/β-catenin signaling has a lower Wnt-pathway activity due to the delayed response, which can be explained by the additional steps required for the formation of the 'signalosome' (complex formation of Wnt-pathway membrane

proteins). The model, thus, proposed two mechanisms to be further experimentally validated. However, due to the absence of experimental techniques to distinguish the different pools of proteins and their cellular localizations, these still remain to be studied and experimentally validated on a fine-grained level. Finally, the model represents an important basis to further study external influences of Wnt/β-catenin signaling.

7.3 Hypothesis of alternative mechanisms: Aurora kinase A signaling

The cell cycle regulator, Aurora kinase A (AURKA), is often gained, as a consequence of chromosomal aberrations, and overexpressed in colorectal adenoma-to-carcinoma progression, but has been shown to further induce Wnt and Ras-MAPK signaling. On a large scale this can be viewed as a positive feedback loop, the induction of which might start in both components. In Chapter 3, we applied a network-aware analysis (Dittrich et al., 2008; May et al., 2016) to this problem, using gene expression data generated upon *AURKA*-knockdown experiments in two CRC cell lines. Here, we described a model of the influence of AURKA on Wnt and Ras-MAPK signaling, finding a set of genes that could play an important role connecting these components.

The Petri net model of Wnt/β-catenin signaling presented in Chapter 2 can be repurposed and extended to study other biological problems; for instance, by combining it with other signaling pathways or by including other factors that have been shown to (or have been hypothesized to) influence further induction of this pathway. AURKA is therefore a perfect candidate to be included into the Wnt/β-catenin Petri net model. The components and interactions connecting AURKA to the Wnt/β-catenin signaling pathway described in the network model in Chapter 3 can be included to the Wnt/β-catenin signaling Petri net model. Challenges here are to determine the model parameters and to identify the directionality of the proposed interactions.

7.4 Exhaustive Petri net modeling: plausible underlying mechanisms of genetic interactions

Genetic interactions between pairs of genes occur when mutating both genes has a different phenotype than mutating the genes individually (Phillips, 2008). Genetic interactions are highly prevalent; a considerable number has been identified to date (Lehner et al., 2006). The molecular mechanisms underlying genetic interactions, however, are not well defined (Large et al., 2017; Lagator et al., 2017; van Leeuwen et al., 2017), but can be studied using network models. Here, parts of the models are based on known factors and parts on hypotheses. By testing combinations of the hypothesized variables mechanisms for the genetic interactions can be predicted. In Chapters 4 and 5 we describe an exhaustive modeling approach to derive

mechanisms for genetic interaction patterns in yeast creating a total of 9,172,034 Petri net models. These models are based on interacting gene pairs and known phenotypes observed between them, but their connectivity, interaction types and interaction strengths are hypothetical. Future studies could include increasing the model size, the granularity of factors included, or applying machine-learning techniques to analysis the wealth of simulation data produced.

7.5 Generation of executable models

Executable models of biological systems are based on formalisms that use computer algorithms to simulate the biological behavior (Fisher and Henzinger, 2007; Fisher et al., 2014). To generate a computational model four inputs need to be defined: *i*) the model, *ii*) parameters for the different conditions, *iii*) simulation rules, and *iv*) analysis rules, as described in detail in Chapter 6. Manual generation of these inputs is a time consuming procedure, and not trivial at all, as we have shown extensively in Chapters 2, 4, and 5 of this thesis. The BioASF framework described in Chapter 6, however, enables automatic generation of executable models based on BioPAX model definitions. Thus, the manual step normally required to define a model is not needed here, because in BioPAX models, the biological knowledge is already readily available in a formal language. The other three inputs: parameters, simulation rules and analysis rules are manually generated, thus, full automatic generation of executable models is not feasible with BioASF. This would first require that these three inputs also are available in a formalized form and that these formalisms can be converted in accordance with the format and granularity of the various computational modeling frameworks. Further, the four specified inputs are not only a good basis for future automatic generation of executable models, but are at the same time providing sufficient information for these models to be reproducible and reusable in other projects.

7.6 Conclusions

The network models presented in this thesis were created in close collaboration with computational and experimental biologists to generate testable hypotheses and to improve the understanding of cellular processes. We described how the models are generated, what is required to make them executable, and that complete descriptions of the modeling process and data will improve model generation.

References

Bonzanni N, Feenstra KA, Fokkink W, Krepska E. What Can Formal Methods Bring to Systems Biology? FM 2009: Formal Methods. Springer, Berlin, Heidelberg; 2009. pp. 16-22.

Bonzanni N, Garg A, Feenstra KA, Schutte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics*. 2013, 29, :i80-8.

Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. *Bioinformatics*. 2009, 25, 2049-56.

Cadigan KM, Peifer M. Wnt signaling from development to disease: insights from model systems. *Cold Spring Harb Perspect Biol*. 2009, 1, a002881.

Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res*. 2012, 22, 398-406.

Clevers H, Nusse R. Wnt/beta-catenin signaling and disease. *Cell*. 2012, 149, 1192-205.

Dittrich MT, Klau GW, Rosenwald A, Dandekar T, Muller T. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics*. 2008, 24, i223-31.

Fisher J, Henzinger TA. Executable cell biology. *Nat Biotechnol*. 2007, 25, 1239-49.

Fisher J, Piterman N, Bodik R. Toward synthesizing executable models in biology. *Front Bioeng Biotechnol*. 2014, 2, 75.

Goodwin BC. Temporal Organization in Cells; a Dynamic Theory of Cellular Control Process. London: Academic Press. 1963.

Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*. 2008, 9, 770-80.

Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theoret Biol*. 1969, 22, 437-67.

Lagator M, Paixao T, Barton NH, Bollback JP, Guet CC. On the mechanistic nature of epistasis in a canonical cis-regulatory element. *Elife*. 2017, 6, e25192.

Large EE, Padmanabhan R, Watkins KL, Campbell RF, Xu W, McGrath PT. Modeling of a negative feedback mechanism explains antagonistic pleiotropy in reproduction in domesticated Caenorhabditis elegans strains. *PLoS Genet*. 2017, 13, e1006769.

Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. *Nat Genet*. 2006, 38, 896-903.

May A, Brandt BW, El-Kebir M, Klau GW, Zaura E, Crielaard W, et al. metaModules identifies key functional subnetworks in microbiome-related disease. *Bioinformatics*. 2016, 32, 1678-85.

Nusse R, Varmus H. Three decades of Wnts: a personal perspective on how a scientific field developed. *EMBO J*. 2012, 31, 2670-84.

Petri CA. Kommunikation mit Automaten. PhD Thesis, Technische Universität Darmstadt, Bonn, Germany. 1962.

Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008, 9, 855-67.

Torkamani A, Schork NJ. Identification of rare cancer driver mutations by network reconstruction. *Genome Res*. 2009, 19, 1570-8.

van Leeuwen J, Pons C, Boone C, Andrews BJ. Mechanisms of suppression: The wiring of genetic resilience. *Bioessays*. 2017, 39, 10.1002/bies.201700042.

Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. *J Comput Biol*. 2011, 18, 507-22.

Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med*. 2004, 10, 789-99.

Acknowledgements

Finally, my thesis is done! I am so grateful for all the people who have supported me while I was writing this book. It would not have been possible without you.

To my (co)supervisors; **Jaap Heringa** and **Anton Feenstra**, thanks for giving me this opportunity. Thanks for all the discussions and all your invaluable feedback, constructive criticism and ideas. **Jaap**, thanks for your guidance both in science and in life. Thanks for bringing me into the world of bioinformatics. I am really happy that you encouraged me to be involved in the ISCB RSG Netherlands and the international Student Council. **Anton**, thanks for always being ready to help and discuss any aspect of my PhD research. The passion you put into these discussions and your creative ideas were extremely valuable. Thanks for keeping me motivated all the way to the end.

Next, I would like to extend my gratitude to the assessment committee; **Martine J. Smit**, **Ioannis Xenarios**, **Lodewyk Wessels**, **Paulien Hogeweg** and **Sanne Abeln**, thanks for your time and effort in reading my thesis. I am honoured to have you in my committee. Your research has been a great inspiration to me.

Obviously, I did not do all of this research on my own. I consider myself lucky to have worked together with many great people in different collaborations. **Martine J. Smit**, **Renée van Amerongen**, **Nika Heijmans** and **Folkert Verkaar**, thank you for introducing me to the world of Wnt signaling. You taught me things I could not have learned by reading papers. I really enjoyed our meetings where we discussed biological experiments and computational simulations of Wnt signaling. **Martine J. Smit**, **Marco Siderius** and **Jeffrey R. van Senten**, it was pleasant working with you and exploring the connections between the HCMV virus and Wnt signaling. Thank you for your enthusiasm and open-minded discussions on possible alternative hypotheses. I am curious to see where this is project is going. **Remond J. A. Fijneman**, **Linda J. W. Bosch**, **Sanne R. Martens-de Kemp** and **Beatriz Carvalho**, I really enjoyed our collaboration. It was intriguing to study Wnt signaling in the perspective of complex tumors. Development of this project already started at the second year of my PhD research. Thank you for all the countless hours going in to this project and also for all general scientific guidance. **Patrick Kemmeren** and **Saman Amini**, it was pleasant working together with you and applying modeling on a new problem: genetic interactions. I really enjoyed all of our discussions on this complex problem.

Most of my PhD research was spent among great colleagues. Past and current IBIVU members; **Ali**, **Bas**, **Cico**, **Erik**, **Gunnar**, **Hannes**, **Jochem**, **Maurits**, **Mohammed**, **Nicola**, **Punto**, **Reza**, **Sanne**, **Ted** and **Qinzheng**, thank you for the time we worked together, it was a pleasure sharing offices with you, discussing science over coffee, spending time together outside work; running, playing football and drinking beers. **Sanne**, thank you for all our discussions leading to two collaborative papers, and thank you very much for looking after me. **Reza**, IBIVU

is really lucky to have you on the team. I really enjoyed being able to discuss science with you every Friday. I was also really pleased to co-author your paper on BioPAX. **Mohammed**, when I moved to the Netherlands you really helped me settle. Thanks for taking me to Koninginnedag, running with me every week and helping me move house (true lifesaver there). I look up to your scientific work and am proud to have been by your side during the defense of your PhD. **Ted**, it was a pleasure sharing the office with you. I really appreciate all our scientific discussions, but also those related to non-scientific things (swimming, cultural differences and what not). It means a lot to me that you introduced me to your family and I hope that you are killing it in Cambridge. **Erik** and **Maurits**, it has been really important having you on my side through my PhD time. It comforted me to feel that we were going through the same thing. I really enjoyed all the dinners and beers we have had together. Apologies to Maurits for serving 'things he has to eat with his hands'. Finally, I would like to thank you for your scientific input on the introduction, discussion, and summary of this thesis.

To all the students I have worked with; **Matthias**, **Kiki**, **Franz**, **Luc**, **Maxim**, **Anne-Merel**, **Mounir**, **Roel**, **Oceane** and **Olga**. Thank you for making me look at this project from unexpected angles and showing me new possibilities that I could not have thought of myself. **Olga**, you started out as a student, but you quickly developed into an invaluable colleague. For a long while Sunday mornings were reserved for Skype calls with you. It is thanks to your hard work and amazing programming skills that we completed two manuscripts, and I am forever grateful for your huge contribution to this thesis. I wish you all the best in your own PhD research. I am without doubt that you will nail it.

To all the people I got to know through ISCB RSG Netherlands; **Agnieszka**, **Julian**, **Roy**, **Saman**, **Sepideh**, **Sjoerd** and **Susanne**, thanks for all the countless hours we spent on discussing topics such as: PhD-retreat organisation, ideas for social games, and logo design. Some topics were more important than others, but through RSG I learned so much about organisation and networking and their importance in science. To the European Student Council Symposium 2016 team, **Kevin**, **Nanne**, **Susanne**, **Farzana**, **Nazeefa**, **Mehedi** and **Bart**, thanks for our collaboration. I am proud to have organized this successful international event together with you. To all, thanks for all the good times, and I look forward to meeting you again.

While finalizing this book I have simultaneously taken the next step in my career. To my colleagues in Leiden; **Barend**, **David**, **Eleni**, **Erik**, **Kees**, **Kristina**, **Marco**, **Mark** and **Rajaram**, thanks for being such sweet colleagues and providing such a nice working atmosphere. This has been really important for me during this period.

Anna og **Theis**, jeg er så glad for, at det var lige præcis var jer to jeg lærte at kende da jeg flyttede til Amsterdam. **Anna**, du var her ikke så længe, men du er et vigtigt minde i min tid i Amsterdam. Jeg husker at du boede i en husbåd på Prinsengracht, og at vi godt kunne lide at gå til koncerter, danse og udforske Amsterdam. Men, jeg husker specielt de dage da du boede hos mig, hvor du spillede guitar og lavede mad

til mig. Tak for de gode tider. **Theis**, tak fordi du har været min ven igennem alle disse år. Jeg er rigtig glad for alle de stunder vi har haft sammen. Takket været dig har jeg haft mere rock and roll i mit liv i Amsterdam. **Jason** and **Simon**, thanks for good times in Amsterdam. Thanks for showing me IJhallen, dinner parties (where you cooked delicious, exotic food!) and restaurants I would never have found myself.

Til mine allerbedste venner i Danmark; **Cecilie**, **Mie**, **Stine**, **Hjalte** og **Ulrik**, tak for alle de gode stunder vi har haft både i København og i Amsterdam i løbet af de sidste par år. Tak for de utallige gange hvor jeg har overnattet hos jer, eller hvor I har arrangeret grillfest, brunch, eller surprise fødselsdag - bare for mig. I er de allerbedste. I var en stor del af mit liv da jeg boede i Danmark, men I er stadig en stor del af mit hjerte. Tak.

Til míni vinfólk í Føroyum (nevnd og ónevnd), eg eri sera takksom fyri at hava tykkum. Takk fyri stuðul og allar góðar løtir. Tað er altíð deiligt at koma heim og hittast við tykkum. **Elisabeth**, **Heidi**, **Annika**, **Halla**, **Joan**, **Anna Maria**, **Jóannes** og **Janus**, takk fyri vitjan í Amsterdam og fyri fantastiskar løtir í Havn, Keypmannahavn og Rom. Eisini takk fyri góðar løtir á Ólavsøku, annað jólakvøld, konsertir, bindiklub, brunsj og fyri góð prát um alt og einki.

Til mína familju; **mammu**, **babba**, **Guri**, **Rannvá**, **Høgna**, **Vár**, **ommu**, **abba** og øll hini. Takk fyri stuðul, kærleika og álit ígjøgnum hesa tíð. Takk fyri vitjan í Amsterdam og takk fyri, at eg altíð kann búgva hjá tykkum, tá ið eg eri í Føroyum ella í Danmark. Eg eri ofta blivin spurd um, hvørjum eg arbeiddi við, men eg havi ikki altíð verið líka góð at svara hesum spurningi. Vónandi gevur hendan bókin eitt sindur av innlit í hesum.

"*One, remember to look up at the stars and not down at your feet. Two, never give up work. Work gives you meaning and purpose and life is empty without it. Three, if you are lucky enough to find love, remember it is rare and don't throw it away.*" - Stephen Hawking

Voor mijn lieve **Robin**, dank je voor je steun en liefde. Dit proefschrift is er ook dankzij jou. Hoeveel heb ik niet geschreven met jou naast mij? Veel van onze vrije tijd is opgegaan aan het schrijven van dit boek en daarom is het ook van jou. Ik bewonder de passie die je hebt voor alles in je leven, het geeft mij heel veel inspiratie; eten en drinken, muziek, spontaniteit, enthousiasme voor dingen die je leuk vindt en liefde voor alle mensen in je omgeving. Bedankt dat ik je familie en vrienden heb leren kennen. Tot slot, bedankt dat je me laat zien dat er in het leven, naast onderzoek, ook nog andere dingen belangrijk zijn.

Summary

Cancer is a complex disease that is characterized by multiple and diverse genetic alterations (mutations), affecting cellular molecules and the functions that they regulate. The most important goal of studying complexes diseases is to better understand how mutations cause disease initiation and development.

The thousands of molecules in a cell are interconnected in different types of networks that work together to regulate the numerous cellular functions and thereby the cellular behaviour. Thus, altered cellular molecules may lead to altered networks, which in turn may cause disease progression. Investigating these altered networks may therefore lead to a better understanding of the molecular cause of complex diseases. However, investigation of these networks is often only possible with computational efforts due to their complexity.

In this thesis, we present our research, where we use computational modeling and analyses to explore altered cellular networks. A computational model of a cellular network can be used to demonstrate and predict how it behaves in a healthy or diseased condition. Additionally, we clearly describe our computational frameworks so they can be easily understood and reused by other researchers.

Colorectal cancer is one of the most common types of cancer, causing 8.5% of the cancer related deaths in the developed world (694,000 per year). In the initial steps of colorectal cancer development, hyperactivated cell growth turns healthy cells into non-cancerous tumor cells due to specific mutations. At a later stage, these cells can transform into cancerous tumor cells and from there on spread. In the initial steps, the Wnt protein plays an important role. In healthy cells, the Wnt protein activates the 'Wnt/β-catenin signaling pathway' leading to accumulation of the β-catenin protein. Specific mutations, however, provide deregulated (misleading) Wnt/β-catenin signaling, in the absence of the Wnt protein, leading to increased β-catenin levels and formation of non-cancerous tumor cells. In Chapter 2, we created a computational model of the Wnt/β-catenin signaling pathway. By simulating healthy and deregulated Wnt/β-catenin signaling with this model, we could explore the differences in their molecular interactions leading to different levels of β-catenin.

In the next step of colorectal cancer progression the non-cancerous tumor cells further progress to cancerous tumor cells due to deregulation of additional signaling pathways such as the 'Ras-MAPK signaling pathway'. In the majority of colorectal cancer this is associated with 'chromosomal aberrations', where large portions of chromosomal DNA are gained (copied) or lost (removed). Aurora kinase A, or AURKA, is a gene that is frequently gained and subsequently overexpressed (the protein product occurs in higher levels than normal) because of these events. In Chapter 3, we investigated the molecular interactions of the effect of AURKA-gain on Wnt/β-catenin and Ras-MAPK signaling in two colorectal cancer cell lines by applying a network analysis. Here, we combine gene expression data with protein-protein interaction data to identify genes that were deregulated by

AURKA. Although most of these genes differed between the two cell lines, genes in the Wnt/β-catenin and Ras-MAPK signaling pathways were enriched in both.

In complex diseases, mutations in multiple genes lead to functional deregulation that cannot be described by the additive effects of each individual mutation. This is due to 'genetic interactions', where mutations in two different genes cause unexpected phenotypes compared to the phenotypes of the individual mutations. In Chapter 4, we studied possible molecular networks for the genetic interaction pattern called inversion, where the phenotype of two combined mutations is the opposite of the two single mutations. To that end, we created all possible computational models of inversion. By analysing these models we observed that a combination of strong and weak interactions is a crucial component for the molecular networks underlying inversion.

In Chapter 5, we described how we generated the computational framework used in Chapter 4 to investigate all the possible genetic interactions models. Here, we described model creation, parameter definition, model simulation, assignment of specific genetic interaction patterns to each model and further analysis of possible molecular interactions for the different genetic interaction patterns. In Chapter 6, we described a different framework that converts descriptive biological models from a database to executable models that can be simulated. Using this framework, amongst other things, we repeated our simulation of Wnt/β-catenin signaling from Chapter 2.

Samenvatting

Kanker is een complexe ziekte die wordt gekenmerkt door een groot aantal verschillende genetische afwijkingen (mutaties), die de moleculen van een cel en hun functies beïnvloeden. Het belangrijkste doel van het onderzoek naar complexe ziektes is om een beter inzicht te verkrijgen in hoe genetische mutaties het startschot voor een ziekte geven, en verdere ontwikkeling hiervan veroorzaken.

De duizenden moleculen binnen een cel zijn verbonden via verschillende soorten netwerken die samen de cel een groot aantal functies verlenen en daarmee de cel zijn gedrag geven. Als moleculen veranderen door een bepaalde mutatie, kan dit hun interactie in een netwerk beïnvloeden, en als gevolg hiervan een verstoring of een ziekte veroorzaken. Onderzoek naar deze verstoorde netwerken geeft ons daarom een beter inzicht in de moleculaire oorzaak van complexe ziektes. De complexiteit van de netwerken maakt dat onderzoek vaak alleen mogelijk is met behulp van computers.

In dit proefschrift presenteren we ons onderzoek waarin we computermodellen en analyses gebruiken om meer te ontdekken over verstoorde cellulaire netwerken. Een computermodel van een cellulair netwerk kan gebruikt worden om te laten zien of te voorspellen hoe het zich gedraagt in een gezonde of een 'zieke' toestand. Verder beschrijven we de manier van werken (het computationele raamwerk) zodanig dat deze gemakkelijk te begrijpen is en hergebruikt kan worden door andere onderzoekers.

Darmkanker is een van de meest voorkomende vormen van kanker, verantwoordelijk voor 8.5% van de kanker-gerelateerde sterfgevallen in de westerse wereld (694.000 per jaar). In de beginfase van darmkanker, ingeleid door bepaalde mutaties, zorgt hyper-geactiveerde celgroei ervoor dat gezonde cellen veranderen in niet-kanker tumorcellen. In een latere fase kunnen deze cellen in kankercellen overgaan, en kunnen dan gaan uitzaaien. Het Wnt eiwit speelt in de beginfase een belangrijke rol. In gezonde cellen activeert het Wnt eiwit de zogenoemde Wnt/β-catenine signaleringsroute, wat leidt tot ophoping van het β-catenine eiwit. Door specifieke mutaties in componenten van dit signaaloverdrachtsnetwerk kan het 'op hol' slaan en zonder dat Wnt eiwitten aan de buitenkant van de cel aangeboden worden toch grote hoeveelheden β-catenine produceren, om zodoende de overgang naar niet-kanker tumorcellen te veroorzaken. In hoofdstuk 2 wordt een computationeel model beschreven, waarmee we normale en verstoorde Wnt/β-catenine signalering hebben gesimuleerd, om zodoende de verschillen in de moleculaire interacties die leiden tot verschillende hoeveelheden van het β-catenine eiwit op te helderen.

In de volgende ontwikkelingsstap van darmkanker, veranderen de niet-kanker tumorcellen in kankercellen, door de deregulatie van andere signaalroutes zoals de 'Ras-MAPK signalering'. In de meerderheid van darmkankers, is dit gekoppeld aan 'chromosomale afwijkingen', waarbij grote delen van het chromosomale DNA

vermeerderd (gekopieerd), of verloren is geraakt. Aurora kinase A, of AURKA, is een gen dat door deze gebeurtenissen vaak vermeerderd is, en vervolgens tot overexpressie komt (het eiwitproduct komt in hogere aantallen voor dan normaal). In hoofdstuk 3, onderzoeken we de moleculaire interacties van het effect van AURKA-vermeerdering op Wnt/β-catenine en Ras-MAPK signalering in twee verschillende darmkanker cellijnen, door middel van netwerk analyse. Hier combineren we gen-expressie data met eiwit-eiwit interactie data, om genen te identificeren die gedereguleerd worden door AURKA. Alhoewel de meeste van de gevonden genen verschillend waren tussen beide cellijnen, vinden we in beide gevallen opvallend veel genen uit de Wnt/β-catenine en Ras-MAPK signaleringspaden.

In complexe ziektes, leiden mutaties in meerdere genen tot functionele deregulatie, die niet beschouwd kan worden als de optelsom van alle afzonderlijke mutaties. Dit komt door 'genetische interacties', waarbij mutaties in twee verschillende genen een onverwacht fenotype opleveren, vergeleken met de fenotypes van de afzonderlijke mutaties. In hoofdstuk 4 bestuderen we alle mogelijke moleculaire netwerken op het voorkomen van een genetisch interactiepatroon dat 'inversie' genoemd wordt, waarbij het fenotype van de twee gecombineerde mutaties het omgekeerde is van de twee enkele mutaties. Hiervoor hebben we alle mogelijke computationele modellen van inversie gemaakt, om te laten zien dat een combinatie van sterke en zwakke interacties een cruciaal element is voor de aan inversie onderliggende genetische netwerken.

In hoofdstuk 5 beschrijven we hoe we het computationele raamwerk gemaakt hebben dat in hoofdstuk 4 gebruikt is om alle mogelijke genetische interactie modellen te onderzoeken. Hier beschrijven we het maken van de modellen, definiëren van de parameters, simuleren van de modellen, en het toekennen van bepaalde genetische interactiepatronen aan elk model, en de verdere analyse van mogelijke moleculaire interacties voor elk van de genetische interactiepatronen. In hoofdstuk 6, tenslotte, beschrijven we een nieuw raamwerk dat beschrijvende biologische modellen uit een database omzet in 'uitvoerbare' (executable) modellen die gesimuleerd kunnen worden. Met behulp van dit raamwerk konden we onder andere onze simulatie van Wnt/β-catenine uit hoofdstuk 2 reproduceren.

Samandráttur

Krabbamein er ein kompleks sjúka, eyðmerkt við nógvum og ymiskum genetiskum broytingum (mutatiónum), ið ávirka kyknumolekylir og funktiónirnar sum tær regulera. Tað týdningarmiklasta endamálið við gransking innan kompleksum sjúkum er at fáa betur innlit í, hvussu mutatiónir elva til byrjan og menning av sjúkum.

Molekylir, í túsundatali í eini kyknu, eru sambundin í ymsum slag av netverkum, ið arbeiða saman fyri at regulera tær mongu kyknufunktióninar og harvið atburðin hjá kyknuni. Broytt kyknumolekyl kunnu føra til broytt netverk, sum hinvegin kunna elva til sjúkumenning. Gransking av hesum broyttu netverkum kann økja okkara vitan um molekyleru orsøkina til kompleksar sjúkur. Tó er gransking av hesum netverkum ofta bara møgulig vid telduroynd, orsaka av teirra kompleksiteti.

Í hesi doktararitgerð bera vit fram okkara gransking, har vit brúka teldumodellir og -greining til at granska broytt kyknunetverk. Eitt teldumodell av einum kyknunetverki kann nýtast til at vísa á og spáa um, hvussu kyknan ber seg at í frískari ella sjúkari støðu. Harafturat greiða vit týðiliga frá okkara teldurammu, so hon lættliga kann fatast og endurnýtast av øðrum granskarum.

Krabbamein í tarminum er eitt av mest vanligu krabbameinsløgunum og er atvoldin til 8,5% av deyðsføllunum, tengd at krabbameini, í ídnaðarlondunum (694.000 um árið). Á byrjunarstigi í tarmkrabbamenningini, syrgir hyperaktiveraður kyknuvøkstur fyri, at frískar kyknir broytast til ikki-krabbameins tumorkyknir, orsaka av serligum mutatiónum. Á seinni stigi kunnu hesar kyknir broytast til krabbameins tumorkyknir, og spjaðast haðani frá. Á byrjunarstigi hevur Wnt proteinið ein týdningamiklan leiklut. Wnt proteinið aktiverar 'Wnt/β-catenin signaleringsrásina' í frískum kyknum, ið førir til upphópan av β-catenin proteininum. Serstakar mutatiónir, í pørtum av hesari signaleringsrás, kunnu tó føra til dereguleraða, villeiðandi signalering, hóast Wnt proteinið ikki er til staðar. Hetta førir til økta β-catenin mongd og skapan av ikki-krabbameins tumorkyknum. Í kapittul 2 skaptu vit eitt teldumodell av Wnt/β-catenin signaleringsrásini. Við at simulera fríska og dereguleraða Wnt/β-catenin signalering við hesum modelli, kundu vit granska munin í teirra molekylera samvirkni, sum førir til ymsar mongdir av β-catenin.

Á næsta stigi í tarmkrabbamenningini, broytast ikki-krabbatumorkyknirnar til krabbakyknir, orsaka av deregulering av øðrum signaleringsrásum, so sum 'Ras-MAPK signaleringsrásini'. Í flestum tarmkrabba førum, hevur hetta samband við kromosomisk avvik, har stórar nøgdir av kromosomiskum DNA er økt (kopiera) ella burturmist. Aurora kinase A, ella AURKA, er eitt gen sum er økt og síðani yvirekspressionera (proteinúrtøkan kemur fyri í størri mongdum enn vanligt) orsaka av hesum tilburðum. Í kapittul 3 granskaðu vit tær molekyleru interaktiónirnar av ávirkanini av AURKA-øking á Wnt/β-catenin og Ras-MAPK signalering í tveimum krabbakyknulinjum, við at nýta netverksgreining. Her sameindu vit genekspressións data við protein-protein interaktión data, fyri at eyðkenna gen sum

var deregulera av AURKA. Hóast flestu av hesum genum vóru ólík ímillum tær báðar kyknulinjurnar, funnu vit í báðum førum ríka gen á Wnt/β-catenin og Ras-MAPK signaleringsrásini.

Í kompleksum sjúkum føra mutatiónir í fleiri genum til funktiónella deregulering, ið ikki kann lýsast við tí additivu effektini av hvørjari einstakari mutatión. Hetta kemur av 'genetiskum interaktiónum', har mutatiónir í tveimum ymsum genum útvega eina óvantaða fenotypu, samanborið við fenotypurnar av hvørjari einstakari mutatión. Í kapittul 4 granska vit møgulig molekyler netverk av tí genetisk interaktiónsmynstrinum, ið verður kalla inversjón, har fenotypan av teimum báðum seimeindu mutatiónunum er øvut av teimum einstaku mutatiónunum. Tískil gjørdu vit øll tey møguligu modellini av inversjón. Við at greina hesi modell, sóu vit, at ein samanseting av sterkum og veikum interaktiónum er ein týdningarmikil partur av tí molekylera netverkinum, grundleggjandi fyri inversjón.

Í kapittul 5, vístu vit, hvussu vit gjørdu teldurammuna, sum var nýtt í kapittul 4, til at granska øll tey møguligu genetisku interaktiónsmodellini. Her lýsa vit, hvussu modellini eru gjørd. Parametrini vóru allýst, modellini vóru simuleraði, áseting av serstøkum genetiskt interaktiónsmynstrum til hvørja modell og framhaldandi greining av møguligum molekylerum interaktiónum av teimum ymsu genetisku interaktiónsmynstrinum. Í kapitul 6, lýstu vit eina aðra teldurammu, sum broytir lýsandi lívfrøðilig modell frá einum databasa til útinnandi modellir, sum kunna verða simuleraðar. Við at nýta hesa teldurammu, m.a. endurtóku vit simuleringina av Wnt/β-catenin frá kapittul 2.

Curriculum vitae

Annika Jacobsen was born on January 14, 1986 in Tórshavn, Faroe Islands. She obtained her Bachelor's degree in 2009 and her Masters' degree in 2012 in Engineering from the Technical University of Denmark. During her studies she completed three research projects on comparative genomics in the group of prof. David W. Ussery one of which resulted in a publication in Microbial Ecology. During her Masters' studies she spent one semester taking bioinformatics courses at Katholieke Universiteit Leuven, Belgium. She did her Masters' thesis at Statens Serum Institut, Copenhagen, studying genetic variation in *E. coli* bacteriophages under the supervision of prof. Karen Angeliki Krogfelt. Annika started her PhD studies in December 2012 in the Integrative Bioinformatics group at the Vrije Universiteit Amsterdam, The Netherlands. Her PhD studies were focussed on exploring the molecular basis of cellular networks using computational modeling and analysis in collaboration with different research groups. During her PhD studies she was the secretary of the ISCB Regional Student Group Netherlands and amongst other chaired the biannual European Student Council Symposium 2016 in The Hague. Since March 2017 Annika has been working in the BioSemantics group in Leiden University Medical Center, The Netherlands, as a postdoctoral researcher on FAIR rare disease data.

Publications

1 **Annika Jacobsen**, Rene S. Hendriksen, Frank M. Aaresturp, David W. Ussery, Carsten Friis. The Salmonella enterica pan-genome. *Microb Ecol.* 2011, 62(3), 487-504. doi: 10.1007/s00248-011-9880-1.

2 Farzana Rahman, Katie Wilkins, **Annika Jacobsen**, Alexander Junge, Esmeralda Vicedo, Dan DeBlasio, Anupama Jigisha, Tomás Di Domenico. Highlights from the tenth ISCB Student Council Symposium 2014. *BMC Bioinform*. 2015, 16, Suppl 2:A1. doi:10.1186/1471-2105-16-S2-A1.

3 **Annika Jacobsen**, Nika Heijmans, Folkert Verkaar, Martine J. Smit, Jaap Heringa*, Renée van Amerongen*, K. Anton Feenstra*. Construction and Experimental Validation of a Petri Net Model of Wnt/β-catenin Signaling. *PLoS One*. 2016, 11(5), e0155743. doi:10.1371/journal.pone.0155743.

4 Reza Haydarlou, **Annika Jacobsen**, Nicola Bonzanni, K. Anton Feenstra, Sanne Abeln, Jaap Heringa. BioASF: A Framework for Automatically Generating Executable Pathway Models Specified in BioPAX. *Bioinformatics*. 2016, 32(12), i60-i69. doi:10.1093/bioinformatics/btw250.

5 Bart Cuypers, **Annika Jacobsen**, Ben Siranosian, Kevin Schwahn, Ashley Mae Conard, Nanne Aben, Mehedi Hassan, Nazeefa Fatima, Susanne M.A. Hermans, Melissa Woghiren, Pieter Meysman, Farzana Rahman, Anupama Jigisha. Highlights from the ISCB Student Council Symposia in 2016. *F1000Res*. 2016, 5, ISCB Comm J-2852. doi:10.12688/f1000research.10389.1

6 **Annika Jacobsen**, Linda J.W. Bosch, Sanne Martens-de Kemp, Beatriz Carvalho, Richard J. Dobson, Emanuele de Rinaldis, Gerrit A. Meijer, Sanne Abeln, Jaap Heringa, Remond J.A. Fijneman, K. Anton Feenstra. Aurora kinase A (AURKA) interaction with Wnt and Ras-MAPK signalling pathways in colorectal cancer. *Sci. Rep*. 2018, 8(7522). doi:10.1038/s41598-018-24982-z

7 Gillian S. Townend*, Friederike Ehrhart*, Henk J. van Kranen, Mark Wilkinson, **Annika Jacobsen**, Marco Roos, Egon L. Willighagen, David van Enckevort, Chris T. Evelo, Leopold M. G. Curfs. MECP2 variation in Rett syndrome - an overview of current coverage of genetic and phenotype data within existing databases. *Hum. Mutat.* 2018, 39(7), 914-924. doi:10.1002/humu.23542

8 Yllka Kodra, Jérôme Weinbach, Manuel Posada-de-la-Paz, Alessio Coi, S. Lydie Lemonnier, David van Enckevort, Marco Roos, **Annika Jacobsen**, Ronald Cornet, S. Faisal Ahmed, Virginie Bros-Facer, Veronica Popa, Marieke Van Meel, Daniel Renault, Rainald von Gizycki, Michele Santoro, Paul Landais, Paola Torreri, Claudio Carta, Deborah Mascalzoni, Sabina Gainotti, Estrella Lopez, Anna Ambrosini, Heimo Müller, Robert Reis, Fabrizio Bianchi, Yaffa R. Rubinstein, Hanns Lochmüller, Domenica Taruscio. Recommendations for Improving the Quality of Rare Disease Registries. *Int J Environ Res Public Health*. 2018, 15(8). pii: E1644. doi: 10.3390/ijerph15081644

9 Saman Amini*, **Annika Jacobsen***, Olga Ivanova, Philip Lijnzaad, Jaap Heringa, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren. The ability of transcription factors to differentially regulate gene expression is a crucial component of the mechanism underlying inversion, a frequently observed genetic interaction pattern. *bioRxiv.* 2018. doi: 10.1101/449520.

10 **Annika Jacobsen***, Olga Ivanova*, Saman Amini*, Patrick Kemmeren, Jaap Heringa, K. Anton Feenstra. A framework for exhaustive modelling of genetic interaction patterns using Petri nets. Submitted.

* Equal contribution